وزارة التعليم العالي
جامعة أم القرى
كلية التربية
قسم علم النفس

مشكلات الدلالة الإحصائية في البحث التربوي وحلول بديلة

إعداد الطالب
عادل بن أحمد بن حسن بابطين

إشراف الأستاذ الدكتور
ربيع بن سعيد بن علي طه

بحث تكميلي لنيل درجة الماجستير في علم النفس تخصص إحصاء وبحوث

الفصل الدراسي الأول ١٤٢٢ هـ

بسم الله الرحمن الرحيم

ملخص الرسالة

هدفت الدراسة إلى الكشف عن أهم مشكلات الدلالة الإحصائية في البحث التربوي، والتعرّف على أهم المفاهيم والأساليب الإحصائية التي يمكن أن تقدم حلولاً لتلك المشكلات، وصياغة أساليب جديدة لحلها، والكشف عن واقع تلك المشكلات في الرسائل العلمية لكلية التربية بجامعة أم القرى، وكذلك الكشف عن واقع وجود تلك المفاهيم والأساليب الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في الرسائل العلمية لكلية التربية بجامعة أم القرى. ولتحقيق أهداف الدراسة صيغت التساؤلات التالية: ما أهم مشكلات الدلالة الإحصائية في البحث التربوي ؟ وما أهم المفاهيم والأساليب الإحصائية التي يمكن أن تقدم حلولاً لتلك المشكلات؟ وما واقع تلك المشكلات في الرسائل العلمية بكلية التربية بجامعة أم القرى؟ وما واقع وجود المفاهيم والأساليب الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في تلك الرسائل العلمية ؟ وقد اعتمدت الدراسة على منهجين بحثيين هما: تحليل المعلومات (الوثائق) للإجابة عن التساؤل الأول والثاني والثالث، والمنهج الوصفي التقويمي للإجابة عن التساؤلين الرابع والخامس، حيث قام الباحث باختيار عينة قصدية من رسائل الماجستير بكلية التربية بجامعة أم القرى من أقسام (الإدارة التربوية والتخطيط، والمناهج وطرق التدريس، وعلم النفس) التي أجيزت خلال الأعوام ١٤١٧، ١٤١٨، ١٤١٩ هـ والتي استخدمت اختبارات الدلالة الإحصائية، حيث بلغ عددها (٩٩) رسالة وبواقع (٧٢٪) من المجتمع الأصلي.

وتوصلت الدراسة إلى أنّ أهم مشكلات اختبارات الدلالة الإحصائية في البحث التربوي هي: استخدام نتائج الدلالة الإحصائية كتفسير لأهمية النتائج، وكتفسير لتأثير المعالجة، واستخدام قيمة الاحتمالية (r) لتقدير حجم التأثير، وكتفسير لاحتمالية الفرضية الصفرية نفسها، ولتقويم إعادة النتيجة، إضافة إلى التحيّز للنتائج الدالة إحصائياً، وصياغة التساؤلات البحثية والفرضيات، والطبيعة الثنائية لنتائج اختبار الدلالة الإحصائية. كما توصلت إلى أنّ أهم المفاهيم والأساليب الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في البحث التربوي هي: تحليل القوة الإحصائية، وتقدير حدود الثقة، وتقديرات حجم التأثير، وتحليل الإعادة. أمّا الأساليب الجديدة التي توصلت إليها الدراسة لحل مشكلات الدلالة الإحصائية في البحث التربوي فهي: إعادة صياغة الأسئلة البحثية والفرضيات بحيث تركز على حجم التأثير وقيمة النتائج بدلاً من مجرد وجود أو عدم وجود الفروق، وفحص البيانات بشكل موسع.

أمّا فيما يخص واقع مشكلات الدلالة الإحصائية في رسائل الماجستير بكلية التربية بجامعة أم القرى فقد كشفت الدراسة أنّ (٩٨٪) من رسائل الماجستير بكلية التربية بجامعة أم القرى تعتمد على نتائج اختبارات الدلالة الإحصائية فقط في تفسير النتائج، وأنّ (٨٥,٩٪) من الرسائل تستخدم نتائج اختبارات الدلالة الإحصائية في تفسير تأثير المعالجات أو العوامل المستقلة، وأنّ (٧,١٪) من الرسائل تستخدم قيم الاحتمالية (r) في تقدير حجم التأثير، كما أنّه لا يوجد أي استخدام لقيمة الاحتمالية (r) في تقدير احتمالية إعادة النتائج، ولا يوجد أي استخدام لقيمة الاحتمالية (r) في تقدير احتمالية حدوث الفرضية الصفرية نفسها، وأنّ نسبة النتائج الدالة إحصائياً (٤٧,٦٪) من العدد الإجمالي لنتائج اختبارات الدلالة الإحصائية حيث لا يوجد تحيّز لتلك النتائج، وأنّ نسبة التساؤلات عن الفروق ذات الدلالة الإحصائية (٦١,٢٪) من مجموع التساؤلات البحثية، وأنّ النتائج التي قيمة الاحتمالية فيها ما بين ٠,٠٥١ و ٠,٠٦ لم تتجاوز نسبة (٠,١٪) من المجموع الكلي للنتائج.

أمّا فيما يخص واقع المفاهيم والأساليب الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في رسائل الماجستير بكلية التربية بجامعة أم القرى فإنّه لا يوجد أي استخدام لتحليل القوة الإحصائية، ولا لتقدير حدود الثقة، ولا لتحليل الإعادة، أمّا مقاييس حجم التأثير فلم تستخدم إلاّ في ٢٪ من الرسائل.

وقد أوصت الدراسة بالتركيز في مقررات الإحصاء الاستدلالي على فلسفة الأساليب الإحصائية ومفاهيمها وحدودها وافتراضاتها ومشكلاتها، و توجيه المقرّرات نحو المزيد من أساليب تقدير المؤشرات مقارنة بأساليب اختبار الفرضيات، وتوسيع خطوات اختبار الفرضيات التقليدية بحيث تتضمن قياس مؤشرات الدلالة العملية في نفس السياق، وتطوير معايير تحكيم الرسائل العلمية بحيث لا تسمح بالاعتماد الوحيد على نتائج اختبارات الدلالة الإحصائية في تفسير النتائج دون إرفاقها بمؤشرات للدلالة العملية مع تفسيرها وكذلك ضرورة تفسير النتائج غير الدالة إحصائياً وبرهنتها بمؤشرات أخرى، وتشجيع إجراء الدراسات المعتمدة على إعادة دراسات سابقة، وتأسيس مركز متخصص للاستشارات البحثية والإحصائية بجامعة أم القرى.

الإهــــــداء

إلــى .. من تاقت نفسي لزوم صحبتها.. أقبّل قدميها.. وأمسح دمعة عينيها.. وأفرح لبسمة شفتيها.. وأكون عبداً صاغراً بين يديها.. أمي الحبيبة.. رعاها الإله.. ومتعنا بها.. وبارك أنفاسها وأيامها..

إلــى .. من رعاني طفلاً صغيراً.. وشجعني راشداً كبيراً.. تعجز ذاكرتي أن تجد له إلاّ الجميل.. والدي الكريم الحليم .. جزاه الله عنّي وأبنائي وإخواني الخير الوافر الجزيل..

ورزقني الله برّهما.. وحسن صحبتهما.. وردّ شيء من جميل عطائهما.. والحظو برضاهما وقبولهما.. وأستغفر الله للتقصير في حقهما..

إلــى .. زوجتي الحبيبة القريبة.. رفيقة دربي.. وشريكة حياتي.. ونبع حنان أبنائي.. وردة متفتحة.. وشمعة مضيئة.. وروح طيبة محبّة.. أثاب الله جميل صنعها..

إلــى .. أبنائي الصغار الكبار.. ثبات .. ندى.. رغد.. أحمد.. وإياد.. رياحين الحياة.. ومذاق الحب والرحمة.. أصلحهم ربّي.. وكتب لهم التوفيق والسعادة.. في الحياة الدنيا والآخرة.. اللهم آمين..

إلــى .. إخواني.. أشقاء وأصدقاء.. من شاركوني طعم الحياة.. وبادلوني الاخلاص والوفاء.. فكانوا لي نعم الرفقاء.. استضيء بآرائهم.. وأتشجع بأقوالهم.. وهم بحمد الله كثر.. ربّي أعلم بهم.. جزاهم الله خير الجزاء..

أهــدي هــذا العمــل المتواضــع..

{ ربنا تقبل منّا إنّك أنت السميع العليم }

شكر وتقدير

الشكر والحمد والثناء لله تعالى على ما وهبنا من النعم.. فقد أحيانا من عدم.. وهدانا من ضلالة.. وعلمنا من جهالة.. وعافانا وآوانا وكسانا.. فلله تعالى الحمد كما ينبغي لجلال وجهه وعظيم سلطانة..

ثمّ الشكر الوافر الجزيل لسعادة المشرف على هذه الدراسة الأستاذ الدكتور/ ربيع بن سعيد طه الذي ساندني ودعمني في جميع خطوات هذه الدراسة، فلم يبخل عليّ بمشورة، ولقد أقلقته أوقات راحته، فلم أجد منه إلاّ الصدر الرحب والخلق الطيب. فجزاه الله تعالى خيراً، وأدام عليه الفضل العظيم.

كما أتقدّم بالشكر الجزيل لسعادة الأستاذ الدكتور/ زايد بن عجير الحارثي الذي قام مشكوراً بتحكيم خطة هذه الدراسة ومناقشتها، والذي كان لتشجيعه الشخصي لي أبلغ الأثر في اتمام هذه الرسالة ، فجزاه الله خيراً، وأتمّ له النعم.

كما أشكر سعادة الدكتور/ عبدالحفيظ بن سعيد مقدّم الذي قام مشكوراً بقبول مناقشة هذه الدراسة، والذي قام بتوجيهي في المراحل المتقدمة من الدراسة، ولقد كان دوماً لي نعم المعين، جزاه الله خير الجزاء. كما أشكر سعادة الدكتور الفاضل/ ثابت بن محمد القحطاني، الذي قام مشكوراً بتحكيم خطة هذه الدراسة والذي غمرني بعنايته، وخصّني بتوجيهاته،والذي لا أنسى دعمه ووقفاته، أدام الله عليه الفضل وأجزل له الأجر.

والشكر موصول مع التقدير والعرفان لرئيس قسم علم النفس السابق، سعادة الدكتور/ محمد بن جعفر جمل الليل، وللرئيس الحالي للقسم سعادة الدكتور/ حسين بن عبدالفتاح الغامدي اللذين وجدنا منهما كل تفهّم وخلق رفيع، واللذين جاهدا لتذليل العقبات من أمامنا فجزاهما الله تعالى خير الجزاء.

كما أتقدّم بالشكر والعرفان والامتنان لجامعة أم القرى، وكلية التربية بالجامعة لما قدمتاه لنا من تسهيلات ورعاية واهتمام، وأخص بالشكر قسم علم النفس وكافة أعضاء هيئته التدريسية الذين تلقينا على أيديهم البيضاء العلوم والتوجيه، فأجزل الله لهم الأجر والثواب.

كما أتقدم بالشكر والتقدير والعرفان لوزارة المعارف، ممثلة بإدارة التعليم بالعاصمة المقدسة التي أتاحت لي فرصة التفرّغ للدراسة بمرحلة الماجستير، وأخص بالشكر سعادة الأستاذ/ محسن بن عبدالرحمن بهادر الذي شجعني ودعمني بكل ما يستطيع جزاه الله تعالى خير الجزاء.

كما أشكر سعادة الدكتور/ صالح بن موسى الضبيبان المدير العام للإشراف التربوي بوزارة المعارف، الذي لم يتوان أبداً في تقديم كل التسهيلات التي تمكنني من إنجاز الرسالة، وسعادة الأستاذ/ خالد بن عبدالعزيز الخريجي مساعد المدير العام للإشراف التربوي بوزارة المعارف بالرياض الذي قدّم لي كل التشجيع والمساعدة والتسهيل خاصة في المراحل النهائية من إعداد الرسالة.

كما أشكر زميلي القدير المشرف العام بالإدارة العامة للإشراف التربوي سعادة الدكتور/ عبدالله بن صالح المقبل والزميل الدكتور/ عزت عبدالحميد اللذين تطوعا مشكورين بمساعدتي في مراجعة مسودة هذه الدراسة، وتحكيم أداتها، واللذين قدما لي العديد من الملاحظات، جزاهما الله تعالى خير الجزاء. والشكر موصول إلى أخي وزميلي سعادة الأستاذ / عبدالله بن عمر بن جحلان الذي أهدى لي العديد من الملاحظات جزاه الله تعالى خير الجزاء.

كما أتقدم بالشكر الجزيل لسعادة المشرف التربوي العام لمادة التربية الوطنية الأخ الأستاذ/ ياسر محمد أمين بخاري على اهتمامه الكبير ومساهمته في تحسين العرض التقديمي الذي قدّم أمام لجنة مناقشة وإجازة هذه الدراسة، أحسن الله له الأجر وأدام عليه الفضل، وسعادة الأستاذ/ محمد بن عمر المهنا وسعادة الأستاذ/ منصور بن ضيف الله العصيمي المشرفين العامين لمادة اللغة الإنجليزية بالوزارة اللذين تفضلا بمساعدتي في ترجمة بعض المصطلحات، جزاهم الله تعالى خير الجزاء.

كما أشكر منسوبي مكتبة الملك فهد الوطنية بالرياض، الذين زودوني مشكورين بجميع القوائم التي تطلبها البحث، وأخص منهم بالشكر الجزيل الأستاذين محمد عبدالرحمن العيسى، وسعد عبدالرحمن الغنّام ، المختصين بإيداع الرسائل الجامعية، حيث عملت معهما على مدى شهرين متتابعين، أتردّد عليهما بشكل يومي، فلمست تعاونهما واهتمامهما جزاهما الله خير الجزاء.

ولن يفوتني أن أشكر والديّ القديرين الغاليين، الذين شاركاني مشواري بالابتهال والدعاء، والمتابعة والسؤال، والترقب والانتظار، جزاهما الله تعالى خير ما يجازي والدين عن ولدهما، كما أشكر شريكة حياتي ، زوجتي الحبيبة الوفية التي كفتني الكثير والكثير لأصرف كل وقتي في إتمام رسالتي، جزاها الله خير الجزاء، وأدام هناءها، وملأ حياتها سعادة، وأعقبها في الآخرة فوزاً وفلاحاً.

كما لن يفوتني أن أشكر إخواني وزملائي: الأستاذ/ نصار الظاهري، والأستاذ/ عبدالحميد عبدالقدوس، والأستاذ/ عبدالعزيز بخش الذين لا أحصي وقفاتهم الجميلة ووفائهم منقطع النظير أجزل الله لهم العطاء والثواب.

وأخيراً وليس آخراً .. أشكر كل من ساهم معي مادياً أو معنوياً، وأعتذر لمن فاتني ذكره ولم أتمكن في هذا المقام من شكره، سائلاً الله تعالى أن لا يضيع لهم أجراً..

وصلى الله على نبينا محمد وآله وصحبه وسلم،،،

المحتويات

الفصل الأول : مشكلة الدراسة

الفصل الثاني : الإطار النظري والدراسات السابقة

الفصل الثالث : إجراءات الدراسة

الفصل الرابع: تحليل النتائج ومناقشتها

أولاً : إجابة الأسئلة الأول والثاني والثالث

ثانياً : إجابة السؤالين الرابع والخامس

الفصل الخامس : الخلاصة والتوصيات

قائمة الجداول

قائمة الرسوم البيانية

قائمة الأشكال

قائمة الملاحق

الفصل الأول

مشكلة الدراسة

مقدمة

الحمد لله رب العالمين. والصلاة والسلام على سيد الأنبياء والمرسلين، نبينا محمد وآله وصحبه أجمعين، وبعـــد.

يعتمد البحث التربوي اختبارات الدلالة الإحصائية كأسلوب إحصائي علمي لتقرير ما إذا كانت الفروق أو الارتباطات أو العلاقات بين المتغيرات دالة إحصائياً. ومع رسوخ هذا الأسلوب أصبحت اختبارات الدلالة الإحصائية لاختبار الفرضيات في الأبحاث التربوية طريقة تقليدية، وكما يقول كيستر Keaster(1998) فقد تحولت اختبارات الدلالة الإحصائية من أسـاليب إلى طقـوس.

يقوم العلم على تراكم المعـرفة، وعلى البناء التنظيمي والهيكـلي لها (Thompson,1994)، وإحدى مشكلات البحث التربوي الخطيرة جداً هي تعارض نتائج الباحثين وتذبذبها (Bangert-Drowns,1991). إنّ هذه المشكلة تشكك في منهجية البحث التربوي، وتقلّل من قيمته العلمية في توليد المعرفة، ومالم يتمكّن البحث التربوي من تحسين استخدامات الأساليب التي تمكّنه من ثبات وصدق نتائج البحث فإنّ هذه المشكلة المنهجية ستستمر.

لقد أصبح استخدام وسوء استخدام اختبارات الدلالة الإحصائية القضية الرئيسة لطرق البحث التربوي الحالي (Kaufman,1998). ولقد كثفت المعارضة للدلالة الإحصائية في السنوات الأخيرة (Burdenski,1999). وارتفعت دعوات علمية إلى ضرورة أن تقاد اختبارات الدلالة الإحصائية إلى مكانها الصحيح في العملية العلمية (Keaster,1998). كما أنّ هناك حاجة علمية إلى أن يتجاوز التركيز في مناقشة قصور الدلالة الإحصائية إلى ما بعد الإدانات الإضافية لتلك الاختبارات، حيث الحاجة إلى مزيد من الاقتراحات لتحسين استخدام ممارسة الاختبارات الإحصائية في اختبار الفروض (Thompson,1997).

تحاول هذه الدراسة الكشف عن أهم مشكلات اختبارات الدلالة الإحصائية، وأهم المفاهيم الإحصائية التي يمكن أن تقدم حلولاً للمشكلات، مع محاولة الوصول إلى أساليب جديدة يمكن اتباعها في البحث التربوي عند محاولة اختبار الفرضيات. كما تحاول الدراسة الكشف عن واقع تلك المشكلات في الرسائل العلمية بكلية التربية بجامعة أم القرى بمكة المكرمة.

مشكلة الدراسة وتساؤلاتها

تاريخياً ساعدت اختبارات الدلالة الإحصائية التقليدية الباحثين في التفكير في معنى بياناتهم، لكن ثبت أنّ الاعتماد عليها بمفردها ينطوي على أخطاء (Huston,1993).خلال عقود ثبت أنّ اختبارات الدلالة الإحصائية غالباً ما يساء استعمالها ،وأنّ تلك الاختبارات محدودة إلى حدما في المساعدة(Thompson,1998b).

فعلى سبيل المثال، أغلب التساؤلات البحثية الأساسية تتضمن أسئلة حول حجم تأثير المتغيرات في المجتمع الإحصائي. لكن اختبارات الدلالة الإحصائية تخفق في الإجابة عن تلك التساؤلات (StatSoft,2000). علماً بأنّ هناك اعتقاداً لدى بعض الباحثين التربويين بأنّ أسئلة البحث أكثر أهمية من فرضياته، وأنّ جزءاً من أسئلة البحث تحتاج لإجابتها إلى أساليب إحصائية أخرى غير اختبارات الدلالة الإحصائية (Hopkins,1997).

لقد تطورت النظرة إلى الدلالة الإحصائية، وأصبح ينظر إليها على أنّها تنتج معلومات قليلة صالحة ومناسبة للتساؤلات البحثية (Shaver,1992)، وأنّ النتيجة ذات الطبيعة الثنائية للدلالة الإحصائية (قبول أو رفض) لم تعد ترضي كثيراً من العلماء الذين يستخدمون الفرضية الصفرية لمعرفة مقدار التأثير أكثر من مجرد التعرّف على وجوده من عدمه (StatSoft,2000). كما تزايد اعتراف الباحثين بقصور ومحدودية اختبارات الدلالة الإحصائية في تشكيل النتائج العلمية (Welge-Crow,1990)، (Palomares,1990). وظهرت معارضة شديدة ومكثفة للدلالة الإحصائية خاصة في السنوات الأخيرة (Burdenski,1999). ومع ذلك، فإنّ استخدام الدلالة الإحصائية وسوء استخدامها يبقيان شائعين، مع أنّه يجب أن يشجع الباحثون على التقليل منها (Shaver,1992).

ويرى بانجيرت -دراونس Bangert-Drowns(1991) بأنّ البحث التربوي في وضعه الحالي يعاني شيئاً من التناقضات، وأنّه يبدو ألاّ أحد يتفق مع اتجاه الآخر، وأنّ الأكثر حزناً أنّه في بعض الأحيان يبدو ألاّ أحد يعرف ما الذي يحدث؟ حتى مع الإعادة المتكررة للتجارب يمكن أن يثبت عدم حسم تعارض النتائج، مما جعل السيناتور فريتز مونداليز يعبر عن ذلك بقوله: إنّ البحث التربوي غالباً ما ينتج نتائج متعارضة (Bangert-Drowns,1991).

هناك اتجاه يركز على نوعية التكامل والشمول التربوي، ويعتبر أحد الاتجاهات الواعدة في البحث التربوي (Bangert-Drowns,1991)، وأنّه كلما كان الباحثون أكثر علماً بفوائد وحدود الدلالة الإحصائية، كانت تفسيرات نتائج البحث أفضل (Huston,1993)، وأنّ هناك ضرورة إلى أن تقاد اختبارات الدلالة الإحصائية إلى مكانها

الصحيح في العملية العلمية (Keaster,1998)، وأنّ العلوم الاجتماعية بشكل عام والتربوية والنفسية بشكل خاص سوف تتقدّم إلى الأمام بشكل أسرع عندما يصبح البحث بحثاً عن التأثيرات الثابتة الواضحة في الحجم في سياق البحث العلمي (Thompson,1997).

يتضح من الخلفية السابقة أنّ هناك تسليماً بالدور التاريخي الذي قامت به اختبارات الدلالة الإحصائية في فهم الباحثين لمعنى بياناتهم، ولكن في السنوات الأخيرة تزايد النقد الشديد للاعتماد على هذه الطريقة التقليدية دون الانتباه إلى قصورها وحدود استخداماتها، ودون البحث عن استراتيجيات جديدة كأسلوب بديل للدلالة الإحصائية في اختبار الفروض.

ومن خلال الاستعراض السابق، يمكن تحديد مشكلة هذه الدراسة كما يلي:

- مع اعتماد البحث التربوي كثيراً على اختبارات الدلالة الإحصائية في تفسير النتائج، إلاّ أنّ مفهوم الدلالة الإحصائية المستخدم في تفسير النتائج يبقى غامضاً ومشكلاً على كثير من الباحثين ممّا يؤدي إلى سوء استخدامه (Keaster,1998).
- عدم إدراك الباحثين لمفهوم الدلالة الإحصائية سبّب ظهور عدّة استخدامات سيئة وخاطئة لاختبارات الدلالة الإحصائية ، ويحتاج البحث التربوي إلى تحديدها (Shaver,1992).
- هناك مشكلات حقيقية علمية ومنهجية أفرزها سوء استخدام اختبارات الدلالة الإحصائية أثرت سلباً في نتائج البحث التربوي (Burdenski,1999).
- اختبارات الدلالة الإحصائية كأسلوب علمي أدت ولا تزال تؤدي دوراً في تفسير نتائج تحليل بيانات البحث التربوي، ولذا يجب صيانتها علمياً من سوء الاستخدام وذلك ببذل المزيد من البحث والتوضيح لحدود استخداماتها وما يمكن وما لا يمكن أن تؤديه في البحث التربوي (Thompson,1995a).
- هناك حاجة للتركيز على مزيد من المفاهيم والأساليب الإحصائية التي يمكن أن تلعب دوراً أكبر في تصحيح الممارسات الخاطئة لسوء استخدام اختبارات الدلالة الإحصائية، أو التي يمكن أن تقدم حلولاً تكميلية أو بديلة لاختبارات الدلالة الإحصائية (Thompson,1997).
- يجب أن يحدد موقع الدلالة الإحصائية في تفسير نتائج البحث التربوي المعاصر بدقة علمية وبدون تحيّز لها أو ضدّها (Keaster,1998).
- هناك حاجة ملحة لتحديد قواعد جديدة يتبعها الباحثون التربويون في اختبار الفرضيات، بحيث تشمل المفاهيم الصحيحة من الطريقة التقليدية والأساليب الجديدة المقترحة، حتى تتبلور أساليب جديدة في اختبار الفرضيات(Kellow,1998) .

وبناءً على ذلك فإنّ هـذه الدراسـة تحاول الإجـابة عـن التسـاؤلات التاليـة:

١. ما أهم مشكلات الدلالة الإحصائية في البحث التربوي؟
٢. ما أهم المفاهيم الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في البحث التربوي؟
٣. ما الأساليب الجديدة لحل مشكلات الدلالة الإحصائية في البحث التربوي؟
٤. ما واقع مشكلات الدلالة الإحصائية في رسائل درجة الماجستير بكلية التربية بجامعة أم القرى بمكة المكرمة.
٥. ما واقع تلك المفاهيم الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في رسائل درجة الماجستير بكلية التربية بجامعة أم القرى بمكة المكرمة.

أهداف الدراسة

تهـدف هذه الدراسـة إلى:

- الكشف عن أهم مشكلات الدلالة الإحصائية في البحث التربوي.
- التعرّف على أهم المفاهيم الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في البحث التربوي.
- صياغة أساليب جديدة لحل مشكلات الدلالة الإحصائية في البحث التربوي.
- الكشف عن واقع مشكلات الدلالة الإحصائية في الرسائل العلمية لكلية التربية بجامعة أم القرى بمكة المكرمة.
- الكشف عن مدى توفر المفاهيم الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في الرسائل العلمية بكلية التربية بجامعة أم القرى بمكة المكرمة.

مسلمات الدراسة

تقوم هذه الدراسة على عدّة مسلمات هـي:

- قدمت اختبارات الدلالة الإحصائية حلولاً للبحث التربوي الكمي خلال الفترة السابقة، ولا يشترط أن يتوقف البحث التربوي عند هـذا الحد.
- الدلالة الإحصائية جوهرية وشائعة في تفسير نتائج البحث التربوي المعاصر وتستحق العناية والبحث والتمحيص.

- هناك حاجة لصيانة البحث التربوي من الممارسات غير الصحيحة عند استخدام نتائج اختبارات الدلالة الإحصائية في البحث التربوي.

أهمية الدراسة

تكمن أهمية هذه الدراسة فيما يلي:

- تبحث هذه الدراسة قضية مثيرة للجدل ومهمة جداً في البحث التربوي، وكما قال كوفمان(١٩٩٨)بأنّ استخدام وسوء استخدام اختبارات الدلالة الإحصائية يعتبران القضية الرئيسة لطرق البحث التربوي الحالي (Kaufman,1998).
- المشكلة التي تعالجها هذه الدراسة استمرت على مدى عدة عقود (Thompson,1998b)، ويمكن أن تستمر ما لم يتوصل البحث التربوي إلى أساليب بديلة.
- هناك تزايد مكثف وكبير لمعارضة الدلالة الإحصائية خاصة في السنوات الأخيرة (Burdenski,1999)، ممّا يزيد أهمية هذه الدراسة، وضرورة استجلاء الصورة الدقيقة للمشكلة، والوصول بالدلالة الإحصائية إلى مكانها المناسب في أدبيات البحث التربوي.
- حتى لو فرضنا بأنّ المعارضات ضد اختبار الدلالة الإحصائية لم تحسم، فعلى حد تعبير دانيل (Daniel(1998b فإنّ التحرك التدريجي نحو الحل يعد مؤشراً إيجابياً.
- تعرض هذه الدراسة المفاهيم العلمية الدقيقة للدلالة الإحصائية وحدود استعمالاتها ممّا يؤثر إيجاباً على تفسير الباحثين لنتائج البحث التربوي بشكل أفضل (Huston,1993).
- المشكلة التي تعالجها هذه الدراسة مشكلة بحثية أساسية على مستوى البحث التربوي العالمي، وعلى مستوى البحث التربوي العربي خاصة.
- تحاول هذه الدراسة ألاّ تتوقف عند مجرد نقد اختبارات الدلالة الإحصائية، بل تتجاوزها للبحث عن اقتراحات جديدة لتحسين ممارسة البحث التربوي.
- وصول هذه الدراسة إلى أساليب جديدة سيؤدي إلى نتائج كبيرة في ميدان البحث التربوي، وكما قال ثومبسون (Thompson(1997 بأنّه يمكن أن تتقدم العلوم الاجتماعية والتربوية والنفسية بشكل أسرع إذا تخلص البحث العلمي في هذه الميادين من بعض مشكلاته الحالية، وإذا أصبح البحث التربوي أكثر ثباتاً في نتائجه.

مصطلحات الدراسة

- المشكلات Problems: تشير المشكلة إلى موقف غامض أو محيّر أو صعب أو ناقص، بحيث تتولد عنها حاجة للوصول إلى الحقيقة، وإشباع النقص، وإزالة العقبات (عبيدات، ١٩٩٩ :ص٩٩).
- الدلالة الإحصائية أو قيمة p (Statistical Significance or p-value): هي احتمالية ملاحظة قيمة الإحصاءة statistic التي تم اختبارها متطرفة عمّا لوحظت عليه على الأقل عندما تكون الفرضية الصفرية صحيحة (Lane,2000).
- البحث التربوي (Educational Research): يعرف فان دالين البحث العلمي بأنّه محاولة دقيقة ومنظمة وناقدة للتوصل إلى حلول لمختلف المشكلات التي تواجهها الإنسانية، وتثير قلق وحيرة الإنسان (عبيدات، ١٩٩٩ :ص٥٢). وحيث أنّ البحث العلمي لا يقتصر على دراسة الظواهر الطبيعية، بل يشمل دراسة الظواهر الاجتماعية والسلوكية، أو الظواهر الإنسانية المتعلقة بمختلف مجالات الحياة الاجتماعية والنفسية والتربوية والاقتصادية (نفس المرجع:ص٥٣)، فإنّ البحث التربوي هو استخدام البحث العلمي في مضمار الميدان التربوي. ويمكن إعادة صياغة تعريف فان دالين ليناسب تعريف البحث التربوي على النحو التالي: هو محاولة دقيقة ومنظمة وناقدة للتوصل إلى حلول للمشكلات التربوية التي تثير قلق وحيرة التربويين.
- الحلول البديلة (Alternative Solutions): هي مجموعة من التوصيات والمقترحات والأساليب التي يمكن صياغتها بأشكال إجرائية بحيث يؤدي اتباعها إلى إزالة مشكلات الدلالة الإحصائية أو تقليصها على أقل تقدير (Lane,2000).
- المجتمع (Population): هو المجموعة الكلية للأعضاء أو المشاهدات أو الدرجات التي تشترك في صفة ما (Lane,2000).
- العينة (Sample): هي مجموعة جزئية من المجتمع مسحوبة بطريقة علمية محدّدة، وذلك لصعوبة فحص كل عضو من أعضاء المجتمع في العادة (Lane,2000).
- توزيع المعاينة Sampling Distribution : هو توزيع الإحصاءات المحسوبة في كل مجموعة من العينات التي حجمها n والتي تكون مسحوبة من نفس المجتمع (McCall 1980:192). وهذا يعني أنّ قيمة الإحصاءة في كل عينة عدد عناصرها n من المجتمع الإحصائي يتم حسابها، ثمّ توزع إحصاءات تلك العينات المحسوبة لنحصل على توزيع المعاينة.

- المتغيّر (Variable): هو خاصية مقاسة أو عشوائية تختلف باختلاف العناصر(Lane,2000).
- المعلم (Parameter): هو كمية عددية تقيس بعض ظواهر درجات المجتمع (مثل المتوسطات، الانحرافات المعيارية، الوسيط،...) (Lane,2000).
- اختبار الفرضيات (Hypotheses Testing): إحدى طرق الإحصاء الاستدلالي، حيث يبدأ الباحث بفرضية حول معلم للمجتمع تسمى الفرضية الصفرية، ثمّ تجمع البيانات ويحدّد مدى قبول الفرضية الصفرية على ضوء تلك البيانات. إذا كانت البيانات مختلفة جداً عن المتوقع عند افتراض صحة الفرضية الصفرية فإنّ الفرضية الصفرية ترفض، أمّا إذا كانت البيانات ليست متباينة عمّا هو متوقع عند افتراض صحة الفرضية الصفرية، فإنّ الفرضية الصفرية لا ترفض، علماً بأنّ الفشل في رفض الفرضية الصفرية لا يعني قبولها (Lane,2000).
- الفرضية الصفرية (Null Hypothesis): هي فرض حول معلم المجتمع، غالباً ما تكون معاكسة لما يعتقد الباحث، توضع ليسمح للبيانات بنقضها، ويرمز للفرضية الصفرية بالرمز H_o (Lane,2000),(McCall,1980).
- الفرضية البديلة (Alternative Hypothesis): في اختبار الفرضيات، الفرضية الصفرية والفرضية البديلة تقدمان بشكل متلازم. إذا رفضت الفرضية الصفرية، قبلت الفرضية البديلة، وإلاّ فلا (Lane,2000; McCall1980)، وهي غالباً ما تدعم توقع الباحث، ويرمز لها بالرمز H_a (Trochim,1999).
- مستوى الدلالة (a) (Significance Level): هو احتمالية الوقوع في الخطأ من النوع الأول (I) ، ويعتبر مستوى الدلالة المعيار المستخدم لرفض الفرضية الصفرية. فعند حساب قيمة الدلالة الإحصائية (p-value) -مع افتراض صحة الفرضية الصفرية - تقارن قيمة (r) بمستوى الدلالة (a)، إذا كانت قيمة (r) أقل من أو تساوي مستوى الدلالة، رفضت الفرضية الصفرية وقيل أنّها دالة إحصائيا وإلاّ فلا، وتقليدياً يستعمل الباحثون إمّا مستوى دلالة ٠,٠٥ أو ٠,٠١ (Lane,2000).
- الخطأ من النوع الأول (I) (Type I error): هو رفض الفرضية الصفرية الصحيحة (Lane,2000).
- الخطأ من النوع الثاني (II) (Type II error): هو الفشل في رفض الفرضية الصفرية الخاطئة (Lane,2000).

- الدلالة العملية (Practical Significance): تهتم الدلالة العملية بمدى كون نتائج البحث مفيدة لعالم الواقع الحقيقي (Kirk,1996). ولذا فإنّ القياسات الإحصائية للدلالة العملية تقرّر إلى أي مدى يمكن تفسير التغيّر الذي يحدث في المتغيّر التابع بواسطة المتغيّر المستقل (Hanson,1987).
- حجم التأثير (Effect Size): هي الدرجة التي توجد فيها الظاهرة في المجتمع، أو مدى تفسير المتغيّر التابع بواسطة المتغيّر المستقل (Kellow, 1998; Nix & Barnette,1998).
- حدود الثقة (Confidence Interval): هي مدى من القيم التي لها احتمالية محدّدة لاحتواء وتضمين المعلم المقدّر (Lane,2000).
- القوة الإحصائية (Statistical Power): هي احتمالية الإصابة في رفض الفرضية الصفرية الخاطئة (Lane,2000).
- تحليل الإعادة (Replication Analysis): هي إجراءات وأساليب إحصائية لتقويم مدى إمكانية إعادة وتكرار النتيجة من خلال تحاليل الإعادة الخارجية (الحقيقية) أو تحاليل الإعادة الداخلية التي تحاكي الإعادة الحقيقية (McLean & Ernet,1998).
- رسائل الماجستير: هي الدراسات والأبحاث العلمية المقدمة من طلاب وطالبات قسم (الإدارة التربوية والتخطيط) وقسم (المناهج وطرق التدريس) وقسم (علم النفس) بكلية التربية كمتطلب تكميلي لنيل درجة الماجستير خلال الأعوام الهجرية (١٤١٧) و (١٤١٨) و (١٤١٩)، والتي استخدمت اختبارات الدلالة الإحصائية.

حدود الدراسة

تقتصر هذه الدراسة على دراسة مشكلات الدلالة الإحصائية في البحث التربوي على افتراض صحّة اختيار الباحث للاختبار المستخدم، وصحة توفر الافتراضات اللازمة لتطبيقه. ولـذا :

- لن تبحث هذه الدراسة أخطاء الباحثين في اختيار الاختبار الإحصائي المناسب عند تحليل البيانات.
- ولن تبحث هذه الدراسة أخطاء الباحثين في عدم التحقق من الافتراضات الخاصة لكل اختبار إحصائي عند تحليل بياناتهم.
- كما تقتصر هذه الدراسة في تعرفها على واقع مشكلات الدلالة الإحصائية على الرسائل العلمية (ماجستير) بكلية التربية بجامعة أم القرى بمكة المكرمة في ثلاثة أقسام هي: الإدارة التربوية والتخطيط، المناهج وطرق التدريس، وعلم النفس، والمجازة خلال الأعوام ١٤١٧هـ، ١٤١٨هـ، ١٤١٩هـ.

الفصل الثاني

الإطـــارالنظـــري

والدراسات السابقة

أولاً - الإطار النظري

تمهيد

تتطلب الدراسة العلمية الارتكاز على إطار علمي دقيق وممحص يعتمد على الأسس والمفاهيم النظرية التي تستند عليها مشكلة الدراسة. وستتناول هذه الدراسة في هذا الإطار المدخل إلى مفهوم الدلالة الإحصائية من خلال الإشارة إلى فلسفة البحث العلمي ودور الإحصاء فيه، ثم الإشارة إلى الإحصاء الاستدلالي وطريقة اختبار الفرضيات. وسيتم التركيز على المفهوم العلمي الدقيق للدلالة الإحصائية، والعوامل المؤثرة في نتائج اختباراتها، وأهم الافتراضات التي ترتكز عليها تلك الاختبارات.

فلسفة البحث العلمي

قبل أن تتبلور الفكرة الحديثة عن البحث العلمي، اعتاد الفلاسفة تسمية البحث العلمي بمصطلح (الاستدلال المنطقي) (Logical Reasoning)، ولذا فمن غير المستغرب أن بعض الاختلافات الأساسية في المنطق انتقلت إلى البحث المعاصر (Trochim,1999). لا يمكن لأي مقدمة منطقية أن تكتمل دون اعتبار بعض الافتراضات الفلسفية. كل الأبحاث تعتمد على افتراضات عن كيفية إدراك العالم، وكيف نستطيع أن نصل إلى أفضل فهم للعالم، وبالطبع لا أحد في الحقيقة يعرف كيف نستطيع أن نصل إلى أفضل فهم للعالم. وقد تجادل الفلاسفة وعلى مدى الفيّتين حول هذه التساؤلات، ويهمنا أن ننظر كيف توجه علماء الاجتماع والعلوم الإنسانية المعاصرون نحو هذه التساؤلات المتعلقة بكيفية إدراك العالم المحيط بنا (Trochim,1999).

أكثر القرارات العلمية تتطلب اختيار بديل من عدّة بدائل ممكنة، بحيث تعمل هذه القرارات بدون معرفة صحة القرار أو خطأه، لأنّها تعتمد على معلومات غير كاملة، لكنّ القرار الرشيد يستخدم إجراءات تؤكد احتمالية و أرجحية نجاحها، ويجب أن تصاغ تلك الإجراءات على هيئة تمكّن أي فرد يستخدم نفس المعلومات أن يصل إلى عمل نفس القرار، ويجب أن ندرك بأنّه لو كانت المعلومات المكتملة متطلبة لعمل القرارات، فإنّ القليل جداً من القرارات ستتخذ، وهذا خلاف الواقع (Stockburger,1998،).

البحث العلمي و الإحصاء

البحث العلمي سواء كان في العلوم الطبيعية أو الإنسانية هو محاولة للإجابة على تساؤلات بشكل منظم وموضوعي ودقيق (McCall,1980). وتتمثّل الخطوات الأولى في البحث العلمي في صياغة التساؤلات بشكل دقيق، والتخطيط للقياسات، واختيار أساليب

منظمة للإجابة على التساؤلات وجمع البيانات، وهذه الخطوات كلها جزء من تصميم التجارب.

لكنّ المهمة لم تنته. فغالباً ماتنتج القياسات أرقاماً كثيرة جداً، ولذا يجب تنظيمها وتلخيصها وربما تمثيلها بيانياً، كما يجب القيام بحسابات للطبيعة العامة للقياسات، والعلاقات التي يجب أن توصف، وأخيراً هناك قرارات يجب أن تتخذ للإجابة على تساؤلات الدراسة، وهذه المهام كلها جزء من الإحصاء.

لذا فإنّ تصميم التجارب والإحصاء مهمان للإجابة عن أغلب تساؤلات العلوم الإنسانية، وغالباً ما تهتم تساؤلات البحث بالعلاقات بين الموضوعات التي تسمى بالمتغيّرات، وعندما يكون التغيّر في موضوع ما (متغيّر ما) ذا علاقة بالتغيّر في موضوع آخر (متغيّر آخر)، فمعنى ذلك أنّ لدينا علاقة سواء كانت تأثيرية سببية أو غير ذلك (McCall,1980).

يتطلب تقدم العلم علاقات جديدة بين المتغيّرات، و تعلمنا فلسفة العلم أنّها لا توجد طريقة أخرى لتمثيل المعنى خلاف شكل العلاقات بين الكميات أو النوعيات أي العلاقات بين المتغيّرات. وغالباً ما يكون الهدف الرئيسي من البحث العلمي خاصة في العلوم الاجتماعية موجه نحو اكتشاف وفهم تلك العلاقات بين المتغيرات خاصة التأثيرات (StatSoft,2000),(Stockburger,1998).

يساعدنا الإحصاء الاستدلالي في تقييم العلاقات بين المتغيرات، ولكن هناك أساسين لأي علاقة بين المتغيرات هما: مقدار العلاقة أو التأثير ، وثبات ومصداقية العلاقة (StatSoft,2000). ولأنّ التساؤلات البحثية لها عدّة أنواع، منها:

- وصفية: وذلك عندما تصمم الدراسة لوصف مايحدث أو مايوجد.
- ارتباطية: عندما تصمم الدراسة للنظر في العلاقات بين متغيرين أو أكثر.
- سببية: عندما تصمم الدراسة لتحديد ما إذا كان أحد المتغيرات أو أكثر يسبّب أو يؤثر على متغيرات أخرى (Lane,2000).

فإنّ الإحصاء يقوم بأدوار كبيرة في الإجابة على تلك التساؤلات، فالتساؤلات الوصفية الأولى يجيب عنها الإحصاء الوصفي، بينما التساؤلان الآخران يجيب عنهما الإحصاء الاستدلالي.

الإحصاء الاستدلالي

يستخدم الإحصاء الاستدلالي للوصول إلى استدلالات حول المجتمع من خلال العينة (Lane,2000)، أي أنّ في الإحصاء الاستدلالي يقوم الباحثون بعمل استدلال وحكم حول ما يوجد على أساس أحداث جزئية فقط، وهذا يتم باستخدام الاحتمالات (McCall,1980).

ولذا يظهر فيه الاستخدام الثاني والكبير لأغراض الإحصاء وهو التعميم من العينة للمجتمع (Hopkins,1997). و تستخدم الجمل الاحتمالية في البحث العلمي لعمل الاستدلالات والأحكام على أساس معلومات جزئية فقط (McCall,1980).

هناك طريقتان تستخدمان في الإحصاء الاستدلالي هما: التقدير واختبار الفرضيات. ففي التقدير، تستخدم العينة لتقدير معلم المجتمع، وتشيّد فيه حدود الثقة حول التقدير. وفي اختبار الفرضيات –شائعة الاستخدام - توضع الفرضية الصفرية في المواجهة ليحدّد الباحث هل البيانات قوية بالقدر الكاف لرفض تلك الفرضية الصفرية أم لا (Lane,2000).

اختبار الفرضيات

اختبار الفرضيات هي إحدى طرق الإحصاء الاستدلالي (Lane,2000). الفرضية هي جملة محدّدة من التوقّع، وهي تصف بلغة محسوسة (أكثر منها نظرية) ما يتوقع الباحث في دراسته. وعندما يفكر الباحثون في الفرضيات، فإنّهم في الحقيقة يفكرون في فرضيتين في نفس الوقت. الفرضية الأولى تصف توقع الباحث، والثانية تصف كل الاحتمالات الأخرى. تسمى الفرضية التي تدعم توقع الباحث الفرضية البديلة H_a ، وتسمى الفرضية التي تصف الخيارات الأخرى والتي غالباً ما تكون معاكسة لتوقع الباحث بالفرضية الصفرية H_o. وحيث أنّ الباحث غالباً ما يحاول أن يبحث عن فرق أو أثر أو علاقة ويحاول إثباتها، فإنّ الفرضية الصفرية غالباً ما تعبّر عن عدم وجود فروق أو أثر أو علاقة، بينما الفرضية البديلة تعبّر عن وجود فروق واختلافات وأثر (Trochim,1999).

إذا كانت توقعات الباحث محدّدة الاتجاه، بمعنى إمّا أن يتوقع بأنّ الأثر إيجابي، أو أن يتوقع بأنّ الأثر سلبي (أحد الخيارين)، فإنّ معنى ذلك أنّ اتجاه توقع الباحث محدّد، وتسمى الفرضية في هذه الحالة أحادية الاتجاه (ذات ذيل واحد)، أي أنّها فرضية موجهة. أمّا إذا كانت توقعات الباحث غير محدّدة الاتجاه، بمعنى أنّه يتوقع أثراً، لكنّه لا يدري هل الأثر إيجابي أم سلبي، فإنّ الفرضية في هذه الحالة تسمى ثنائية الاتجاه (ذات ذيلين)، أي أنّها فرضية غير موجهة (Trochim,1999).

الفرضيتان الصفرية والبديلة مكملتان لبعضهما، وتستوعبان معاً كل الإمكانات المحتملة، وفي نهاية اختبار الفرضيات، يجب بالضرورة ترجيح إحدى الفرضيتين على الأخرى (Trochim,1999). علماً بأنّ ليس كل الدراسات تمتلك فروضا، فبعض الدراسات تكون استكشافية استقرائية غرضها الاستكشاف أكثر من محاولة تطوير بعض الفرضيات المحدّدة أو التوقعات التي يمكن أن تختبر في الأبحاث المستقبلية (Lane,2000).

خطوات اختبار الفرضيات

عملية اختبار الفرضيات تمر بعدّة خطوات هي (Lane,2000):

١. تحديد الفرضية الصفرية H_0 والفرضية البديلة H_a ، وغالباً ما تكون الفرضية الصفرية تنص على عدم وجود فروق بين المجموعات، بينما تكون الفرضية البديلة تنص على وجود فروق (إذا كانت ثنائية الاتجاه)، ويمكن أن تحدّد اتجاهها (إذا كانت أحادية الاتجاه).

٢. تحديد مستوى الدلالة، وغالباً ما تكون ٠,٠٥ أو ٠,٠١ ، حيث يكون مستوى الدلالة الإحصائية هو المعيار المستخدم لرفض أو عدم رفض الفرضية الصفرية (McCall,1980).

٣. حساب الإحصاءة التي تناظر المعلم المحدّد في الفرضية الصفرية. فإذا كانت الفرضية الصفرية تعرّف المعلم بأنّه $m_1 - m_2$ ، فإنّ الإحصاءة التي تحسب هي $M_1 - M_2$.

٤. حساب قيمة الاحتمالية (r) التي هي احتمالية الحصول على إحصاءة محسوبة من البيانات مختلفة عن المعلم المحدّد في الفرضية الصفرية، حيث تعمل الحسابات بافتراض صحة الفرضية الصفرية.

٥. مقارنة قيمة الاحتمالية المحسوبة في الخطوة (٤) بمستوى الدلالة المختار في الخطوة (٢). فإذا كانت الاحتمالية أقل من أو تساوي مستوى الدلالة فإنّ الفرضية الصفرية ترفض، أماّ إذا كانت الاحتمالية أكبر من مستوى الدلالة فإنّ الفرضية الصفرية لا ترفض.

٦. تكون النتيجة دالة إحصائياً إذا رفضت الفرضية الصفرية لصالح الفرضية البديلة.

٧. وصف النتيجة والاستنتاج الإحصائي بطريقة مفهومة.

يتضح ممّا سبق أنّ قرار رفض الفرضية الصفرية أو عدمه (الدلالة الإحصائية للنتيجة أو عدمها) يعتمد على القيمة المحسوبة للاحتمالية (r) ومقارنتها بمستوى الدلالة (a). وعليه فإنّ فهم الباحث لتفسير نتائج بحثه سيعتمد بشكل كبير على مدى فهمه وإدراكه الدقيق والعلمي لمفهوم الدلالة الإحصائية بشكل عام.

مفهوم الدلالة الإحصائية

يرى ثومبسون Thompson(1994) بأنّ قليلاً جداً من الباحثين يفهمون ما تفعله اختبارات الدلالة الإحصائية وما لا تفعله، ولذلك فإنّ نتائجهم عادة ما تكون سيئة التفسير، وحتى إذا فهم بعض الباحثين العناصر المكونة لاختبارات الدلالة الإحصائية، فإنّ مفهومهم عن الدلالة الإحصائية غالباً لا يتكامل في البحث، فعلى سبيل المثال، تأثير حجم العينة على الدلالة الإحصائية ربما يكون معروفاً من قبل الباحث، لكنّ هذا الاستبصار لا يظهر عندما تفسّر نتائج الدراسة التي تشمل عدّة آلاف من أفراد العينة.

وقد أكد لان Lane(2000) هذه المشكلة عندما رأى أنّ هناك خلطاً في المفهوم الدقيق للاحتمالية المحسوبة ($r_{calculated}$) في اختبارات الدلالة الإحصائية (). ويرى هوستون Huston(1993) أنّه كلما كان الباحثون والمحررون وطلاب الدراسات العليا والمجتمع أكثر إحاطة بفوائد وحدود الدلالة الإحصائية ... كان من المرجح، أن يكونوا أكثر قدرة على عمل تفسيرات متقنة ومحنكة ومفيدة لمعنى نتائج البحـث.

تتضمن خطوات اختبار الفرضيات، أن يفترض الباحث بأن تكون الفرضية الصفرية H_o صحيحة، ثمّ يحسب الفرق بين الإحصاءة المحسوبة في العينة والمعلم المحدّد في الفرضية الصفرية H_o . بعد ذلك يحسب الباحث احتمالية ($r_{calculated}$) الحصول على فرق بنفس ذلك المقدار أو أكبر. وتكون هذه القيمة الاحتمالية المحسوبة ($r_{calculated}$) هي احتمالية الحصول على بيانات مختلفة بنفس المقدار عن البيانات الحالية أو بمقدار أكثر اختلافا وذلك مع افتراض صحة الفرضية الصفـرية H_o (Lane,2000).

إذاً في اختبارات الدلالة الإحصائية قيمتان احتماليتان: الاحتمالية الأولى (a) وتسمى أيضاً ($r_{critical}$)، وهي احتمالية اتخاذ القرار الخاطئ في رفض الفرضية الصفرية عندما تكون الفرضية الصفرية في الواقع صحيحة في المجتمع، وهي الحد الأعلى الذي يختاره الباحث لنفسه لارتكاب الخطأ من النوع (I)، ولذا فإنّها من الأساس توضع منخفضة، وعلى أساس ذاتي (Thompson,1994). أمّا الاحتمالية الثانية فهي الاحتمالية المحسوبة ($r_{calculated}$)، وتتراوح كأي قيمة احتمالية من صفر إلى واحد.

وتحسب الاحتمالية المحسوبة ($r_{calculated}$) في سياق عدّة افتراضات حتى تحكم الحسابات بشكل دقيق وتكون للمشكلة الواحدة إجابة واحدة فقط. فمثلاً احتمالية الحصول على متوسطي درجات اختبار ذكاء IQ (٩٩) و (١٠٠) في عينتين، تعتمد على عدّة عوامل من ضمنها المعالم الحقيقية للمجتمعات التي سحبت منها تلك العينات. فمثلاً إذا قارن باحث الاحتمالية المحسوبة ($r_{calculated}$) إذا كان متوسطا المجتمعين بالتوالي: (٩٩) و (١٠١)،

مع الاحتمالية المحسوبة ($r_{calculated}$) فيما لوكان متوسطا المجتمعين (١٠٠) و (١٠٠) فإنّ الاحتمالية المحسوبة ($r_{calculated}$) في الحالة الأولى سوف تكون أعلى، وحيث أنّ المعالم الحقيقية للمجتمع غالباً ما تكون غير معروفة، فإنّه يجب أن يفترض الباحثون قيماً لتلك المعالم. في اختبارات الدلالة الإحصائية دائماً ما يفترض الباحثون أنّ المعالم محدّدة عن طريق الفرضية الصفرية التي يفترض أن تكون صحيحة في هذه الحسابات (Thompson,1994).

هناك عوامل أخرى تؤثر في حساب قيمة الاحتمالية المحسوبة ($r_{calculated}$)مثل حجوم العينات، وتكون الإحصاءات المحسوبة أقل تمثيلاً للمجتمعات كلما صغرت أحجام العينة، وتكون الاختلافات بين إحصاءات العينة أقل احتمالية كلما كبرت حجوم العينات. فمثلاً، إذا كانت العينتان اللتان حجم كل واحدة منهما (٥) لهما متوسطان (٩٠) و (١٠٠)، والعينتان اللتان حجم كل واحدة منهما (٥٠) لهما نفس المتوسطين (٩٠) و (١٠٠)، فإنّ اختبار احتمالية عدم وجود الفروق سوف ينتج قيمة ($r_{calculated}$) منخفضة للعينتين الكبيرتين مقارنة بالعينتين الصغيرتين، وذلك بسبب الافتراض الأصلي للفرضية الصفرية بأنّها صحيحة تماماً (أي أنّ الفروق تماماً معدومة) (Thompson,1994).

إذاً الاحتمالية المحسوبة ($r_{calculated}$) تجيب على التساؤل: بافتراض أنّ بيانات العينة أتت من مجتمع الفرضية الصفرية H_o فيه تماماً صحيحة، ما هي احتمالية الحصول على إحصاءة عينة الباحث عند حجم العينة المعطى؟ بمعنى آخر، هل الاحتمالية المحسوبة ($r_{calculated}$) لنتائج عينة الباحث أصغر من الحد المقبول (a)؟ عندما تكون ($r_{calculated}$) أصغر من ($r_{critical}$) (a) فإنّ قرار الباحث هو رفض الفرضية الصفرية H_o ، وبذلك يعبّر الباحث عن النتيجة بأنّها (دالة إحصائياً) . كل ما يعنيه هذا القرار، أنّ الباحث يعتقد بأنّ نتائج العينة غير محتملة نسبياً، وذلك حسب الافتراضات المعطاة والتي تتضمن الافتراض بأنّ الفرضية الصفرية H_o صحيحة تماماً (Thompson,1994).

وقد عبّر كارفر Carver(1978) عن مفهوم الدلالة الإحصائية بقوله: "ببساطة تعني الدلالة الإحصائية الندرة... النتيجة من وجهة النظر الإحصائية تكون دالة إحصائياً لأنّها تحدث نادراً جداً في العينة العشوائية تحت ظروف الفرضية الصفرية H_o... بذاتها تعني الدلالة الإحصائيــة القـليل أو لا شـــيء" (p.383). وهي بذلك تسمح فقط للباحث بعمل قرار (نعم/لا) فيما يتعلق برفض أو عدم رفض الفرضية الصفرية (Keaster,1988).

وبذلك يكون المفهوم المبسط لاختبار الدلالة الإحصائية بأنّه إجراء ضمن اختبار الفرضيات لتحديد احتمالية النتيجة ($r_{calculated}$) عند افتراض صحة الفرضيـة الصفرية

H_o ، وافتراض عشوائية البيانات (عشوائية الاختيار والتعيين)، وعند حجم عينة n (Kellow,1998),(Shaver,1992). وهذا يعني كما يرى شيفر Shaver(1992) :

- نتيجة اختبار الدلالة الإحصائية هي جملة احتمالية، يعبّر عنها عادة على هيئة ثنائية (dichotomy) لما إذا كانت الاحتمالية أصغر أو أكبر من مستوى (a) .
- الاختبار يعتمد على افتراض أنّ الفرضية الصفرية صحيحة. ولذا فإنّ توزيعات المعاينة النظرية التي تقارن بالتوزيعات الأخرى (التوزيع الطبيعي، توزيعات t، توزيعات F، توزيعات c^2) تحدث بافتراض أنّ المعاينة من المجتمع أو المجتمعات التي تكون فيها الفرضية الصفرية صحيحة.
- العشوائية هي افتراض ضروري لأساس استخدام اختبارات الدلالة الإحصائية.
- حجم العينة له اعتبار حاسم بسبب أنّ الدلالة الإحصائية للنتيجة تعتمد على عدد الحالات المستخدمة.

يتضح ممّا سبق أنّ اختبارات الدلالة الإحصائية تدلّ الباحث على احتمالية حدوث النتيجة فقط، وذلك بافتراض صحة الفرضية الصفرية H_o ، وعند عدد n من حالات العينة العشوائية.

العوامل المؤثرة في نتائج اختبارات الدلالة الإحصائية

معرفة العوامل المؤثرة في قيمة الاحتمالية المحسوبة ($r_{calculated}$) تساعد الباحث في فهم أدق لمفهوم نتائج اختبارات الدلالة الإحصائية. وقد حدّد شنيدر ودارسي سبعة عوامل محدّدة لنتيجة اختبار الدلالة الإحصائية هي:

١ - عدد الحالات المستعملة في الدراسة (حجم العينة).

٢ - مستوى الدلالة المختار.

٣ - التشتت في المتغيّرات ذات العلاقة .

٤ - القوة الفعلية للتأثير.

٥ - مناسبة القياس الإحصائي والاختبارات الإحصائية المستخدمة.

٦ - الفرضية المختبرة.

٧ - تعقيد التحليل (Schneider and Darcy,1984).

وهذه العوامل مفصلة على النحو التالي:

أولاً - تأثير حجم العينة على نتيجة اختبارات الدلالة الإحصائية

يرى ثومبسون Thompson(1989)، وويلكنسون Wilkinson(1992)، وسنايدر ولاوسون Snyder and Lawson(1992) بأنّ أكبر مؤثر على نتيجة اختبار الدلالة الإحصائية من العوامل المذكورة هو عدد الحالات المستعملة في الدراسة أي حجم العينة n ، وأنّ الوصول إلى الدلالة الإحصائية وإن كانت دالة لسبعة خصائص في الدراسة على الأقل إلاّ أنّ حجم العينة هو المؤثر الأساس الذي يجعل النتائج دالة أو غير دالة إحصائياً.

منذ ستين عاماً لاحظ بيركسون أثناء تحليله لما يقارب من (٢٠٠,٠٠٠) عنصر باستخدام اختبارات c^2 للدلالة الإحصائية تأثّر النتائج بحجم العينات وقال:

> الملاحظ الإحصائي الذي لديه خبرة معتبرة في تطبيق اختبار c^2 بشكل متكرّر، سوف يوافقني على نص إفادتي بأنّ: موضوع الملاحظة، عندما تكون الأعداد في البيانات كبيرة نسبياً، فإنّ قيم (r) تميل إلى أن تأتي صغيرة... وبغض النظر كم هي صغيرة التعارضات بين المنحنى الطبيعي والمنحنى الحقيقي للمشاهدات، فإن (r) لاختبار c^2 سوف تكون صغيرة إذا كانت العينة لديها عدد كبير بالقدر الكاف من أعداد المشاهدات... إذا كنّا إذاً نعرف مسبقاً الـ (r) التي ستنتج من تطبيق اختبار c^2 لعينة كبيرة، فإنّه يبدو أنّ لا فائدة من عملها مع عينات أصغر. ولأنّ نتيجة الاختبار معروفة مسبقاً، فإنّه ليس باختبار على الإطلاق (في: **Wilkinson,1992,p.9**).

وقد عرض كيلو Kellow(1998) مثالاً يوضح تأثير حجم العينة في نتيجة اختبار الدلالة الإحصائية كما في الجدول التالي:

جدول ١ : تأثير حجم العينة في نتيجة اختبارات الدلالة الإحصائية

المصدر	SS	r^2	df	M	F_{calc}	F_{crti}	القرار
ن = ١٦							
المفسّر	٣٣٣,٣٣	٠,٣٣٣	٢	١٦٦,٦٧	٣,٢٥	٣,٨	عدم الرفض
غير المفسّر	٦٦٦,٦٧		١٣	٥١,٢٨			
المجموع	١٠٠٠		١٥	٦٦,٦٧			
ن = ١٧							
المفسّر	٣٣٣,٣٣	٠,٣٣٣	٢	١٦٦,٦٧	٣,٥	٣,٧٤	عدم الرفض
غير المفسّر	٦٦٦,٦٧		١٤	٤٧,٦٢			
المجموع	١٠٠٠		١٦	٦٢,٥			
ن = ١٨							
المفسّر	٣٣٣,٣٣	٠,٣٣٣	٢	١٦٦,٦٧	٣,٧٥	٣,٦٨	رفض
غير المفسّر	٦٦٦,٦٧		١٥	٤٤,٤٤			
المجموع	١٠٠٠		١٧	٥٨,٨٢			

المصدر	SS	r^2	df	M	F_{calc}	F_{crti}	القرار
ن = ١٩ المفسّر غير المفسّر المجموع	٣٣٣,٣٣ ٦٦٦,٦٧ ١٠٠٠	٠,٣٣٣	٢ ١٦ ١٨	١٦٦,٦٧ ٤١,٦٧ ٥٥,٥٦	٤	٣,٦٣	رفض

نتائج الجدول السابق تنطبق على تحليل التباين ANOVA باتجاه واحد مشتملاً على مجموعة ثلاثة متوسطات. أعمدة الجدول السابق تمثل: (١) مصدر التباين المفسّر وغير المفسّر(Source)، (٢) مجموع مربعات المكونات المفسّرة وغير المفسّرة (SS)، (٣) مقياس نسبة التباين المفسّر (r^2)، (٤) درجـات الحرية (df)، (٥) متوسط مربعـات لكل مكون للتباين (SS/df) أي (M)، (٦) نسبة F المحسـوبة (MSexplained/MSunexplaine)، (٧) قيمة F الحرجة والتي تعتبر دالة إحصائياً عند مستوى a = ٠,٠٥، (٨) القرار المعتمد على مقارنة F المحسوبة بـ F الحرجة.

من الجدول السابق، يتضح أنّه مع ثبات مقدار التباين المحسوب (r^2) عند قيمة ٣٣٪، فإنّ إضافة عنصر واحد فقط من N=17 إلى N=18 ينتج عنه رفض الفرضية الصفرية والانتقال إلى الفرضية البديلة. نتائج اختبارات الدلالة الإحصائية في المثال السابق تمّ الحصول عليها عند تحديد مستوى الدلالة a = ٠,٠٥ ، وهذا يعني أنّ أي قيمة محسوبة لـ (r) تقع تحت هذا المستوى يمكن اعتبارها دالة إحصائياً عند مستوى ٠,٠٥ ، علماً بأنّ مستوى الدلالة (a) توضع بشكل أكثر حرّية وذاتية، ولذا فلو حسبت قيم (r) عند مستوى دلالة ٠,٠١ لأصبحت جميع القيم السابقة في الجدول غير دالة إحصائياً (Kellow,1998).

وقد عرض هوستون Huston(1993) دراسة لثومبسون عام ١٩٨٩م توضح جدولاً مشابهاً للجدول السابق مع اختلاف أحجام العينات، حيث كانت أحجام العينات (٤، ١٣، ٢٣، ٣٣) عند ثبات حجم التأثير (٣٣,٦٪):

جدول ٢ : مثال آخر لتأثير حجم العينة في نتيجة اختبارات الدلالة الإحصائية

المصدر	SS	r^2	df	M	F_{calc}	F_{crti}	القرار
ن = ٤ المفسّر غير المفسّر المجموع	٣٣٧,٢ ٦٦٥,١ ١٠٠٢,٣	٠,٣٣٦	٢ ١ ٣	١٦٨,٦٠٠ ٦٦٥,١٠٠ ٣٣٤,١٠٠	٣,٢٥	٣,٨	عدم الرفض

المصدر	SS	r^2	df	M	F_{calc}	F_{crti}	القرار
ن = ١٣							
المفسّر	٣٣٧,٢	٠,٣٣٦	٢	١٦٨,٦٠٠	٣,٥	٣,٧٤	عدم الرفض
غير المفسّر	٦٦٥,١		١٠	٦٦,٥١٠			
المجموع	١٠٠٢,٣		٢٠	٨٣,٥٢٥			
ن = ٢٣							
المفسّر	٣٣٧,٢	٠,٣٣٦	٢	١٦٨,٦٠٠	٣,٧٥	٣,٦٨	رفض
غير المفسّر	٦٦٥,١		٢٠	٣٣,٢٥٥			
المجموع	١٠٠٢,٣		٢٢	٤٥,٥٥٩			
ن = ٣٣							
المفسّر	٣٣٧,٢	٠,٣٣٦	٢	١٦٨,٦٠٠	٤	٣,٦٣	رفض
غير المفسّر	٦٦٥,١		٣٠	٢٢,١٧٠			
المجموع	١٠٠٢,٣		٣٢	٣١,٣٢٢			

يتضح من الجدول (٢) أنّ قيمة F أصبحت دالة إحصائياً عندما كان حجم العينة بين ١٣ و ٢٣ عنصراً، كما يتضح أنّ الباحث الذي لا يفهم بشكل دقيق الدلالة الإحصائية سوف يفسر النتائج لحجم التأثير (٣٣,٦٪) بشكل مختلف عندما تكون عدد الحالات في العينة ١٣ مقابل عدد الحالات ٢٣ على الرغم من أنّ حجم التأثير في الجدول ثابت (Huston,1993).

الطريقة الثانية لتوضيح أثر حجم العينات في نتائج اختبارات الدلالة الإحصائية عرضها دانييل (Daniel(1998a على النحو التالي:

بافتراض أنّ باحثاً يجري تجربة تربوية والتي عيّن فيها مجموعتين بطريقتين تدريسيتين مختلفتين، وبعد ذلك أراد أن يقوّم نتائج مقياس الإنجاز. هذا الباحث ربّما يستخدم تحليل التباين لتقويم نتيجة التجربة، حيث يفترض أولاً الفرضية الصفرية H_o التي تنص على عدم وجود فروق بين الأشخاص في نفس الظروف التجريبية، بعد ذلك يحسب قيمة الإحصاءة F والتي عن طريقها سيقوّم الفرضية الصفرية H_o.

وحيث أنّ F ببساطة عبارة عن إحصاءة نسبة، تعتمد على حاصل قسمة متوسط مربعات التأثيرات MS ، مقسومة على متوسط مربعات الخطأ. ونظراً لأنّ متوسط المربعات هي نتيجة حاصل قسمة مجموع المربعات لكل تأثير على درجات الحرية، وأنّ متوسط مربعات

الخطأ سوف يصبح أصغر كلما زاد حجم العينة، فإنّ هذا المقام القاسم لمتوسط مربعات التأثير سينتج قيمة أكبر للإحصاءة F.

وكمثال على ذلك، فإنّ العينة المتكوّنة من ٣٠٠٢ حالة ستكون أكثر احتمالية بمقدار خمس مرّات لانتاج الدلالة الإحصائية من العينة المكونة من ٦٠٢ حالة فقط. وهذا يعني أنّه مع عينة كبيرة مثل ٣٠٠٢ ، حتى الفروق الضيئلة بين المجموعتين ستكون دالة إحصائياً.

وقد لخص ثومبسون Thompson هذا بقوله:

> قيم (r) المحسوبة في دراسة معطاة، عبارة عن دالة لعدّة خصائص للدراسة، ولكنّها متأثرة بالتحديد بخليط التأثيرات لحجم عينة الدراسة وحجوم التأثير للدراسة. وبسبب أن قيم (r) مربكة، نجد أنّه نظرياً يمكن أن تكون لـ (١٠٠) دراسة بأحجام عينات مختلفة و (١٠٠) حجوم تأثير مختلفة قيمة (r) محسوبة واحدة. كما يمكن أن يكون لـ (١٠٠) دراسة بحجم تأثير واحد (١٠٠) قيمة محسوبة ومختلفة لـ (r) **(Snyder&Thompson98,p.7)**.

إنّ الاعتماد الواضح والكبير لنتائج اختبارات الدلالة الإحصائية على حجم العينة n ، جعل العديد من الباحثين مثل ثومبسون Thompson(1997) يذهب إلى أنّه افتراضياً عندما نتعامل مع عينات كبيرة فإنّ كل الفرضيات الصفرية سوف ترفض لأنّ الفرضية الصفرية لعدم وجود فروق في الغالب ليست تماماً صحيحة في المجتمع. وعبّر عن ذلك أيضاً هايز بقوله: "افتراضياً، أي دراسة يمكن أن تظهر نتائج دالة إحصائياً إذا استخدم الباحث مفردات كافية" (Palomers,1990,p.2).

وقد ذهب إلى ذلك أيضاً كوفرسميد عام ١٩٨٨ بقوله " إنّ من النادر أن يكون الارتباط أو الفرق بين المجموعات بالتمام صفراً (كما تفترض الفرضية الصفرية)، ولذا فإنّ الفرضية الصفرية H_o أو فرضية عدم وجود الفرق نادراً ما تبقى. هذا يعني أنّه عملياً مع حجم عينة كبير، أي فرضية صفرية يمكن أن ترفض" (في: Wilkinson92,p.3). وقد قرّر هذا أيضاً ماكلين McLain(1997) بقوله:

> اختبار الدلالة الإحصائية انتقد بشكل متكرّر بسبب اعتماده المفرط على حجم العينة، وعدم القدرة على توضيح معنى النتائج... والباحثون الذين لديهم خبرة العمل مع عينات كبيرة، يدركون حالاً بأنّ كل الفرضيات الصفرية سوف ترفض عند مستوى معيّن من حجم العينة... وذلك لأنّ الفرضية الصفرية المأخوذة حرفياً، دائماً

خاطئة (مرفوضة)... واقعياً، الدلالة الإحصائية يمكن تحقيقها عند حجم معيّن من العينة، إنّها ببساطة تصبح مسألة حصول على عدد كاف من العناصر **(p.7)**

وكما رأى دانييل Daniel(1997) بأنّ الباحث في الحقيقة يمكنه أن يعرف حتى قبل البدء بالتجربة بأنّ اختبار الدلالة الإحصائية هو اختبار لمدى ما إذا كانت العينة كبيرة أم لا. وإذا جمعت بيانات كافية، فإنّ الفرضية الصفرية H_o عادة سوف ترفض، وستكون الفروق الصغيرة في تقديرات معالم المجتمع من العينات الكبيرة، بغض النظر عن عدم دلالتها العملية، تنتج نتائج دالة إحصائياً (Nix and Barnette,1998).

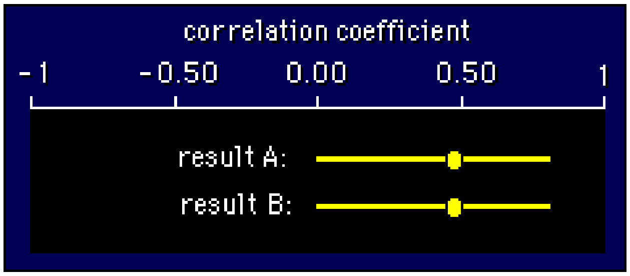
ثانياً : مستوى الدلالة الإحصائية (a)

اختبار الدلالة الإحصائية ينفذ بعد تحديد مستوى الدلالة (a) . مستوى (a) هي احتمالية أن تؤخذ النتيجة المعطاة بسبب أخطاء المعاينة، بمعنى احتمالية ارتكاب الخطأ من النوع (I)، وهو رفض الفرضية الصفرية وهي صحيحة. مستويات الدلالة (a) التقليدية ٠,٠٥ أو ٠,٠١ المستعملة من قبل الباحثين هي تقليد عشوائي(Kellow,1998).

وقد ناقش هوبكينز Hopkins(1997) مفهوم مستويات الدلالة الإحصائية وتطورها بقوله: ما هو الخاص في القيمة ٠,٠٥ ؟ في الحقيقة لاشيء. شخص ما قرّر أنّ هذه القيمة معقولة، ولذا فإنّنا الآن جامدون عندها. (r) < ٠,٠١ أصبحت أيضاً تقليداً لتحديد الدلالة الإحصائية. كلا القيمتين من الآثار القديمة للأيام التي سبقت استخدام الحاسب الآلي، عندما كان من الصعب حساب القيمة الدقيقة لـ (r) لإحصاءة الاختبار، وبدلاً من ذلك استخدم النّاس الجداول لقيم إحصاءة الاختبار المقابلة لقيم قليلة ومختارة وعشوائية لـ (r) وهي ٠,٠٥ ، ٠,٠١ وفي بعض الأوقات ٠,٠٠١ . هذه القيم أصبحت مقدّسة كعتبة حدية للقيم التي تحدّد الدلالة الإحصائية، وعادة ما يكون على الباحث أن ينص على أي واحدة يستخدم. فعلى سبيل المثال، إذا نصّ الباحث على أنّ مستوى الدلالة هو ٥٪ ، إذاً يسمح له باعتبار أي نتيجة لها قيمة (r) تساوي أو أقل من ٠,٠٥ بأنّها دالة إحصائياً.

وقد عرض هوبكينز Hopkins(1997) مثالاً توضيحياً لمأزق هذا التقليد الذي يشبه التقديس لاستخدام قيم محدّدة لمستويات الدلالة حيث رأى بأنّ الباحثين والإحصائيين يدّعون أنّ القرار يجب أن يعمل لمعرفة ما إذا كانت النتيجة دالة إحصائياً. وفقاً لهذا المنطق، إذا كانت r < ٠,٠٥ فالباحث لديه نتيجة قابلة للنشر [لاحظ مشكلة التحيّز في البحث التربوي للنتائج الدالة إحصائيا]، وإذا كانت r > ٠,٠٥ فنتيجة الباحث ليست كذلك.

الشكل (١) يوضح كما يرى هوبكينز Hopkins(1997) حماقة هذه النظرة للعالم، حيث أنّ إحدى النتيجتين دالة إحصـائياً (r < ٠,٠٥)، والأخرى ليست دالة إحصائياً (r > ٠,٠٥)، أيهما صالحة للنشر؟ الإجابة: كلاهما، لأنّ القيمتين الدقيقتين لـ (r) هما (٠,٠٤٩) و (٠,٠٥١) (Hopkins,1997).



شكل ١ : مثال على مأزق الاعتماد على مستوى الدلالة الإحصائية

وكما رأى ماكلين McLane(1997) بأنّ مستوى الدلالة (٠,٠٥) أصبح حارس بوابة عشوائي للمعرفة العلمية، بحيث أنّ النتائج التي لا تقابل المعيار ينظر على أنّها غير مهمة، وأنّ هذه الآلية في تحديد مستوى الدلالة الإحصائية ثم اختبار الدلالة الإحصائية للوصول إلى القرار الإحصائي تبدو محدّدة لصنع القرار ممّا يعطي شعوراً بالموضوعية العلمية، لكنّها في الواقع ليست ذات معنى، لأنّ تكلفة هذه الموضوعية هي المعنى.

ويؤكد هذا التطور في الحسابات نيكس وبارنيت حيث يريان أنّه مع تطوّر وسهولة استخدام برامج الحاسب الآلي للتحاليل الإحصائية، لم يعد الباحث يستعمل الجداول والأساليب اليدوية لاختبارات الدلالة الإحصائية، فبدلاً من ذلك، تقوم الحزم الإحصائية الآلية (برامج الحاسب الآلي) بتزويد الباحث بقيمة (r) التي تستعمل لتحديد رفض أو الفشل في رفض الفرضية الصفرية، حيث تقارن قيمة (r) بقيمة (a) ، ومع ذلك لا تزال مستويات (a) ٠,٠١ ، ٠,٠٥ ، ٠,٠٠١ هي المعيار الحاسم في الحكم على النتائج، وقد كتب كوهين (١٩٩٤) يصف هذا الوضع بقوله "بعد عقود من النقد الحاد، لا تزال تتواصل طقوس اختبار الدلالة الإحصائية الميكانيكية ذات القرار الثنائي للفرضية الصفرية وفقاً للمعيار المقدّس (٠,٠٥) ... أجادل بأنّ اختبارات الدلالة الإحصائية للفرضية الصفرية لم تفشل فقط في دعم تقدّم علم النفس كعلم، بل أعاقته بشكل خطير " (McLane,1997,pp.3-4)

يرى كثير من الباحثين ومنهم ماكلين McLain(1997) أنّ فيشر Fisher دعم مستوى الدلالة (٠,٠٥) كمعيار للدلالة الإحصائية، لكن من المثير الملاحظة أنّ فيشر لم يضع حداً ثابتاً لمستوى الدلالة، بل أنّه دعا الباحثين لتقويم الحالات الفردية بالرجوع للبيانات والنظرية.

ثالثاً : التشتت في المتغيّرات ذات العلاقة

تعتمد نتيجة اختبار الدلالة الإحصائية على قوة الاختبار في رفض الفرضية الصفرية بطريقة صحيحة والحصول على نتائج دالة إحصائياً (Wilkinson,1992), (Huston,1993). هذه القوة الإحصائية للاختبارات بذاتها تعتمد من ضمن ما تعتمد عليه على ثبات البيانات Reliability، الذي بدوره يعتمد على التشتت أو تغاير البيانات، وقد لاحظ العديد مثل هوستون Huston(1993)، ونيكس وبارنت Nix and Barnette(1998) أنّ أكثر طريقة مشاعة لزيادة الثبات والقوة الإحصائية هي زيادة حجم العينة. وعلى أي حال، فإنّ عامل حجم العينة الحاسم قد سبق مناقشته، ولكن يلاحظ أنّ درجة تشتت الدرجات عند ثبات حجم العينة تؤثر في قيمة الاحتمالية المحسوبة ($r_{calculated}$) بحيث أنّ زيادة تقلّب وتغاير وتشتت الدرجات ينتج قيمة أعلى للاحتمالية المحسوبة ($r_{calculated}$)، بينما يؤدي انخفاض درجة التشتت إلى انخفاض قيمة ($r_{calculated}$) ومن ثمّ الحصول على نتائج دالة إحصائياً.

الطريقة الأخرى التي يلاحظ بها تأثير تشتت الدرجات هي قيمة بعض إحصاءات اختبارات الدلالة الإحصائية مثل اختبار F ، حيث أنّ قيمة الإحصاءة هي:

$$F = \frac{S^2_{between}}{S^2_{within}}$$

ونظراً لأنّ المقام (S^2_{within})متأثر باختلافات ضمن المجموعات الذي يعتمد على التباين في الدرجات ضمن المجموعات، فإنّه كلّما قل ذلك التباين انخفضت قيمة المقام وبذلك ترتفع قيمة ($F_{calculated}$) ممّا يزيد احتمالية الحصول على نتائج دالة إحصائياً (Nix and Barnette,1998).

رابعاً : القوة الفعلية للتأثير

في الواقع هذا ما يفترضه الباحث أصلاً، وهو أنّ الفروق التي بين المجموعات سببها تأثير المعالجات. ويبدو أنّ الباحثين ليست لديهم مشكلة في تقرير اعتماد نتائج اختبارات الدلالة الإحصائية على القوة الفعلية للتأثير. المشكلة هي العكس، حيث أنّ الباحثين يفرطون في الاعتقاد بجدارة اختبارات الدلالة الإحصائية في ملاحظة تأثير المعالجات. ولذا فإنّ توضيح تأثير هذا العامل على قيمة اختبارات الدلالة الإحصائية لا يعني هذه الدراسة مثل توضيح العوامل الأخرى التي قد تؤثر فعلاً في نتائج اختبارات الدلالة الإحصائية دون أن يدرك بعض الباحثين تلك المؤثرات في نتائجهم، ممّا يجعلهم يفرطون في الاعتقاد بفائدة تلك الاختبارات والاعتماد عليها في تفسير النتائج ولو كان بشكل غير صحيح (Nix and Barnette,1998).

خامساً : مناسبة القياس الإحصائي والاختبارات الإحصائية المستخدمة

هذا العامل خارج حدود هذه الدراسة، لأنّ هذه الدراسة تبحث في مشكلات الدلالة الإحصائية بافتراض أنّ الباحثين يستخدمون الاختبار الأنسب والذي يستوفي كل الافتراضات اللازمة لتطبيقه.

سادساً : الفرضية المختبرة

وهذا العامل أيضاً خارج حدود هذه الدراسة لأنّه يعتمد على تصميم البحث نفسه، ولأنّ هذه الدراسة التي تبحث عن مشكلات الدلالة الإحصائية تفترض أنّ الباحث يصيغ فرضياته بشكل صحيح وأيضاً يستعمل الاختبار الإحصائي المناسب لها.

سابعاً : تعقيد التحليل الإحصائي

هذا العامل أيضاً يخرج عن حدود هذه الدراسة لأنّه يعتمد على تصميم البحث والذي تفترض هذه الدراسة أنّ الباحث يقوم به بشكل مناسب.

يتضح ممّا سبق أنّ هناك عوامل حاسمة تؤثر في نتائج اختبارات الدلالة الإحصائية وقيم ($r_{calculated}$)، سواء أدركها الباحثون أو لم يدركوها أهمّها حجم العينة المستخدمة ومستوى الدلالة الذي يختاره الباحث وتباين درجات العينة وغيرها من الأسباب. كما يتضح أنّ مناسبة القياس الإحصائي والاختبارات الإحصائية المستخدمة عامل مؤثر في دقة نتائج تلك الاختبارات، علماً بأنّ هذه المشكلة الخطيرة جداً ليست في حدود هذه الدراسة. ونظراً لأنّ هناك افتراضات عامة لاستخدام جميع اختبارات الدلالة الإحصائية بغض النظر عن نوعها، وحتى تكتمل صورة ومفهوم اختبارات الدلالة الإحصائية، قام الباحث بمناقشة أهم افتراضات اختبارات الدلالة الإحصائية وهو العشوائية.

أهم افتراضات اختبار الدلالة الإحصائية

يتطلب نموذج الدلالة الإحصائية البدء بمعاينة عشوائية، ولن ينتج اختبار الدلالة الإحصائية المستخدم بدون عشوائية معلومات صادقة حول احتمالية النتيجة عند صحة الفرضية الصفرية، حيث يرى وينر وآخرون أنّ انتهاك افتراض عشوائية المعاينة للعناصر من المجتمع وعشوائية التعيين للعناصر في المعالجات ربما يمنع الصدق كلياً في الدراسة وذلك بسبب أنّ العشوائية تزودنا بالضمان أنّ الأخطاء موزعة بشكل مستقل بين وفي ظروف المعالجة، كما أنّ العشوائية هي الآلية التي يزال بها التحيّز من ظروف المعالجة " (Shaver,1992). وقد أشار إلى ذلك أيضاً جلاس وهوبكينز (١٩٨٤) بقولهما إنّ الإحصاء الاستدلالي يعتمد على افتراض عشوائية المعاينة من المجتمعات ، وأنّ العينات العشوائية واحدة من أسس بناء اختبار الفرضيات (Shaver,1992).

العشوائية مهمّة لأنّها تساعد في التأكد من استقلالية المشاهدات (أي الأخطاء). وبالإضافة إلى أنّ العشوائية ضرورية لاختبارات الدلالة الإحصائية المعروفة، فإنّ العشوائية (مثل الخطأ العشوائي) هي أساس توزيعات المعاينة مقابل التي تقارن بها النتائج. فعلى سبيل المثال، استخدام توزيع (t) للإجابة على التساؤل (ما احتمالية النتيجة المحدّدة، عند صحة الفرضية الصفرية؟) لن ينتج جملة احتمالية ذات قيمة أو معنى إذا لم تكن العينة عشوائية (Shaver,1992).

إنّ ضرورة وإلزامية العشوائية يمكن أن تبرهن بشكل أكبر عندما يصاغ التساؤل التالي في اختبار الدلالة الإحصائية: ما مدى تمثيل العينة (إحصاءة العينة) للمجتمع (معلم المجتمع) كما هو محدّد في الفرضية الصفرية؟ بدون العشوائية لا يمكن إجابة هذا السؤال بصدق باستخدام اختبارات الدلالة الإحصائية الشائعة (Shaver,1992). وكما أوضح ذلك جلاس وهوبكينز بأنّ طريقة الاختيار العشوائي للعينات يؤكد -مع هامش معروف ومؤكد من الخطأ -تمثيل العينات، ويسمح بتأسيس حدود والتي من المتوقع أن تقع المعالم فيها باحتمالية معينة (Shaver,1992).

إنّ القدرة على تقدير درجة الخطأ بسبب الصدفة (خطأ المعاينة) خاصية مهمة للعينة العشوائية... وإنّه من غير الممكن تقدير الخطأ في عينة صدفية وفي العديد من استراتيجيات المعاينات الأخرى بسبب أنّها تحتوي على أنواع ودرجات غير معروفة من الصدفة وخطأ المعاينة، ولتوضيح ذلك ضرب شيفر Shaver(1992) المثال التالي:

افترض أنّ رجلاً سار إلى غرفة ورأى (١٠) عملات على طاولة، (٨) منها كانت على وجهها الأمامي و (٢) كانت على جانبها الخلفي. تساءل الرجل عمّا إذا كانت تلك العملات متحيّزة! وتساءل عن احتمالية أن يكون هذا الترتيب قد حدث بالصدفة. تناول الرجل مرجعاً فوجد أنّ احتمالية الحصول على (٨) واجهة أمامية من (١٠) هي $\frac{45}{1024}$ أي ٠,٠٤٤، ولأنّ هذه الاحتمالية أقل من مستوى (a) التقليدية ٠,٠٥، فإنّه استنتج أنّ النتيجة من غير المحتمل أن تحدث بالصدفة عند صحة الفرضية الصفرية (٥٠٪ - ٥٠٪)، ولذا فقد اعتبر أنّ لديه برهاناً بأنّ العملات متحيّزة. في الحقيقة مثل هذا الرجل ليس لديه إثبات على تحيّز تلك المجموعة من العملات، والسبب ببساطة أنّه لم يتأكد بأنّ ذلك الوضع للعملات أتى بعد رميها، أي أنّه لم يعرف العملية التي من خلالها وصلت تلك العملات إلى ذلك الوضع. وعليه فإنّ التوزيع النظري الثنائي سيكون قليل التعلّق بالموضوع لأنّ البيانات لم تنتج بالطريقة التي تتوافق مع الافتراض الأساسي لاستخدام التوزيع. تماماً مثل التوزيع الثنائي الذي لم يتمكن من تزويدنا بمعلومات حول التحيّز الممكن في العملات المشاهدة، فإنّ الباحثين التربويين الذين يستعملون مجموعات غير عشوائية لا يمكنهم الحصول على معلومات صادقة حول

احتمالية فروق المجموعات -عند صحة الفرضية الصفرية - باستخدام الاختبارات الإحصائية المعروفة. فالمعاينة العشوائية المتكررة وكذلك التعيين العشوائي ينتجان توزيعات معاينة معروفة، أمّا المعاينة غير العشوائية فإنّها لا تنتج ذلك. وكذلك مقارنة العينة غير العشوائية بتوزيع المعاينة المولدة عشوائياً لا يمكن أن تزودنا بجملة احتمالية صادقة للحدث .(Shaver,1992)

لقد وصف شيفر Shaver(1992) انتهاك هذا الافتراض في أرض الواقع من قبل الباحثين حيث رأى بأنّه في البحث التربوي عادة ما يذهب الباحث إلى مدرسة أو عدّة مدارس ويحصل على مجموعات موجودة (بدون عشوائية المعاينة، ولا عشوائية الاختيار)، ثمّ يجمع البيانات -أحياناً بمعالجة أو أحياناً بدون معالجة - وبعد ذلك يطبق اختبارات الدلالة الإحصائية. نتائج تلك الاختبارات الاستدلالية ليس لها بالضرورة معنى، إلاّ إذا كان أحدهم مهتماً بالمقارنة بمعيار منتقى للاحتمالية مثل الموضح بالسؤال: ما هي احتمالية الحصول على النتيجة (المحصّلة) إذا كانت العينات العشوائية سحبت فعلياً؟ وباختصار، فإنّ اختبارات الدلالة الإحصائية الشائعة تزوّد الباحث بمعلومات محدّدة عن مقدار احتمالية النتيجة بافتراض أنّ الفرضية الصفرية صحيحة وعند عشوائية المعاينة والتعيين وحجم العينة n ، وبدون العشوائية تكون نتيجة اختبار الدلالة الإحصائية عديمة المعنى، وفي أحسن أحوالها، تكون صلتها بجملة احتمالية غير محدّدة أو معروفة (Shaver,1992).

ثانياً – الدراسات السابقة

تمهيد

نظراً لما يتمتع به موضوع الدراسة من استخدامات واسعة وأهمية في البحث التربوي، فقد وجد العديد من الدراسات التي لها صلة بموضوع هذه الدراسة. وقد توزعت تلك الدراسات السابقة بين دراسات عربية وأخرى أجنبية على النحو التالي:

(أ) الدراسات العربية

(١) "الدلالة الإحصائية والدلالة العملية لاختبار (ت) و (ف)"

هذه الدراسة عبارة عن رسالة ماجستير غير منشورة من إعداد الصائغ(١٤١٥). استهدفت الدراسة معرفة واقع الدلالة العملية المصاحبة للدلالة الإحصائية في الرسائل العلمية بكلية التربية بجامعة أم القرى والتي استخدمت اختباري (F) و (t) . كما استهدفت الباحثة تحديد ما إذا كانت هناك علاقة بين الدلالة العملية وحجم العينة، ومستوى الدلالة، ونوع البحث (تربوي أو نفسي).

وقد راجعت الباحثة في دراستها الرسائل العلمية (١٠٣ رسالة) التي استخدمت اختباري (F) أو (t)، ثمّ قامت بحساب قيمتي (h^2) و (Ω^2) كمؤشرين للدلالة العملية ومن ثمّ قامت بمقارنة قيم الدلالة الإحصائية (وعددها ٦٤٨ قيمة) بقيم (h^2) و(Ω^2). كما قامت باختبار الفروق في قيم الدلالة العملية حسب أحجام العينات، ومستوى الدلالة، ونوع البحث (تربوي أو نفسي).

وقد وجدت الباحثة أنّ ٤٧٪ من الفروق الدالة إحصائياً كانت قيمتا (h^2) و (Ω^2) فيها منخفضتين، وأنّ ليس هناك أثر لمستوى الدلالة الإحصائية على الدلالة العملية، وأنه لا توجد علاقة بين الدلالة العملية ونوع البحث (تربوي أو نفسي).

وقد توصلت الباحثة إلى أنّ الكتب العربية تكاد تفتقر للإشارة إلى الدلالة العملية وأهميتها وأساليبها الإحصائية.

(٢) "تقويم استخدامات اختبار كاي تربيع في رسائل الماجستير بكلية التربية بجامعة أم القرى"

هدفت هذه الدراسة التي أعدتها نور (١٤١٣) إلى تقويم استخدامات اختبار (c^{2}) في رسائل الماجستير بكلية التربية بجامعة أم القرى، والتعريف بمصادر الخطأ التي تقلّل من جودة الاستخدام، ومعرفة واقع قوة الاختبار وحجم العينة المستخدم بناء على حجم التأثير لمتغيرات الدراسة.

وقد وجدت الدراسة أنّ الدلالة العملية مرتفعة في اختبارات (c^{2}) وبنسبة (٧٠,١٨٪) من مجموع اختبارات (c^{2}) المستخدمة. كما وجدت الدراسة أيضاً أنّ قوة الاختبار المصاحبة لاختبارات (c^{2}) مرتفعة وبنسبة ٩٦٪ من مجموع استخدامات اختبار (c^{2}) حيث أنّ مقدار القوة الإحصائية في تلك الاختبارات تجاوزت (٠,٥٠). كما وجدت الدراسة أنّ حجم العينة المستخدم في تلك الاختبارات كانت كبيرة ومتضخمة.

(٣) "دراسة تقويمية مقارنة للأساليب الإحصائية التي استخدمت في تحليل البيانات في رسائل الماجستير في كل من كلية التربية بجامعة أم القرى بمكة المكرمة وكلية التربية بجامعة الملك سعود بالرياض"

قام الباحث النجار(١٤١١) بإعداد هذه الدراسة التي هدفت إلى تقويم الأساليب الإحصائية المستخدمة في رسائل الماجستير في كلية التربية بجامعة أم القرى وكلية التربية بجامعة الملك سعود، والتعرّف على الأساليب الإحصائية شائعة الاستخدام في تلك الرسائل. كما قام الباحث بحساب قيمة الدلالة العملية لاختبارات الدلالة الإحصائية المستخدمة.

شملت عينة هذه الدراسة (١٧٧) رسالة جامعية منها (١٢٧) رسالة ما جستير من جامعة أم القرى حتى عام ١٤٠٩هـ، و (٥٠) رسالة ماجستير من كلية التربية بجامعة الملك سعود في نفس الفترة.

وقد وجد الباحث في هذه الدراسة أنّ الأساليب الإحصائية الأكثر شيوعاً في عينة الدراسة هي: (c^{2}) و (F) و (t) و (r) ، وأنّ الدلالة العملية منخفضة في اختبارات الدلالة الإحصائية وبنسبة (٢٧,١٪) من مجموع اختبارات الدلالة الإحصائية المستخدمة.

(٤) "الدلالة العملية وحجم العينة المصاحبتين للدلالة الإحصائية لاختبار (t) في البحث التربوي والنفسي العربي"

هدفت هذه الدراسة التي أعدّها الصياد(١٩٨٨) إلى التعرف على واقع الدلالة العملية من خلال استخدامات اختبار (t) في بحوث رسائل الماجستير والدكتوراة للطلاب والمحترفين (وهم الأساتذة الجامعيون الحاصلون على درجة الدكتوراة والذين يعملون في حقل التدريس والبحث الأكاديمي). كما هدفت الدراسة إلى معرفة حجم العينة ومستوى الدلالة الشائعة في اختبار (t) .

وقد شملت عينة الدراسة (١٣) دورية عربية خلال الفترة ١٩٧٧– ١٩٨٣ م، إضافة إلى رسائل الماجستير المجازة من قبل كلية التربية بجامعة الأزهر خلال الفترة من ١٩٧٣ وحتى عام ١٩٨١م. وقد توصل الباحث إلى أنّ الدلالة العملية في حاجة للإهتمام بها من قبل الباحثين المحترفين بنسبة (٨٤,٩٪) بينما الحاجة للإهتمام بالدلالة العملية من قبل الباحثين الطلاب فهي بنسبة (٨٥,١٪)، وأنّ القيمة الأكثر شيوعاً لمستوى الدلالة الإحصائية هي: (٠,٠١) ثم (٠,٠٥) وأخيراً (٠,١٠). كما وجد الباحث أن أكثر من ٥٠٪ من عينة الدراسة ذات دلالة عملية وسط حسب معيار كوهين، وأنّ العلاقة بين قيمة مربع إيتا وحجم العينة علاقة عكسية، وأنّ الباحثين يتخفون وراء حجم عينة كبير للحصول على الدلالة الإحصائية.

(٥) "جداول تحديد حجم العينة في البحث السلوكي"

قام الباحث الصياد (١٩٨٩) في هذه الدراسة بإيجاد معادلات لتحديد حجم العينة لاختبارات الدلالة الإحصائية الشائعة، كما قام بوضعها في جداول يسهل على الباحث استخدامها حسب الاختبار الإحصائي المستخدم. كما قام الباحث أيضاً في هذه الدراسة بحساب الدلالة العملية (حجم التأثير) كعامل مؤثر في حجم العينة ومتأثر به أيضاً. كما وضع الباحث جداول ارشادية لاختيار الأساليب الإحصائية المناسبة.

توصل الباحث في هذه الدراسة إلى أنّ هناك أزمة في البحث التربوي العربي والتي تكمن في استخدام بعض الاختبارات الإحصائية بجمود دون غيرها، ودون محاولة للتفكير بما هو مناسب وما هو غير مناسب. كما توصل إلى التأكيد على ضرورة الاهتمام بتحليل القوة الإحصائية للاختبارات.

(ب) الدراسات الأجنبية

تميّزت الدراسات الأجنبية السابقة التي لها صلة بموضوع هذه الدراسة بتعدّدها وتنوعها مقارنة بالدراسات العربية. ولذا فقد قام الباحث بتصنيفها إلى صنفين: الأول دراسات تطبيقية تتضمن جمع بيانات من عينات محدّدة وتحليلها إحصائياً، والثاني دراسات غير تطبيقية تعتمد على البحث النظري البحت في جوانب ذات علاقة بموضوع الدراسة. وقد قام الباحث بترتيب جميع الدراسات في كل صنف من الصنفين المذكورين حسب حداثة تاريخ النشر.

أولاً – الدراسات التطبيقية

(١) "النمو التاريخي لاختبار الدلالة الإحصائية في علم النفس – وموقعه في المستقبل"

قام الباحثان هوبارد وريان Hubbard and Ryan(2000) بدراسة النمو التاريخي لشعبية اختبار الدلالة الإحصائية، حيث فحصت عينة عشوائية لأعداد (٢١) مجلة من مجلات الجمعية النفسية الأمريكية American Psychological Association (APA). أنتجت هذه العينة (٨٠٠١) مقالاً إمبريقياً، حيث كانت المقالات التي استخدمت اختبارات الدلالة الإحصائية ٦٥٨٩ أي ما نسبته(٨٢,٤٪). وقد لاحظت الدراسة أنّ الفترة ١٩٤٠ –١٩٥٠ هي الفترة التي حدثت فيها ثورة استدلالية في علم النفس، حيث بدأ فيها انتشار واسع للدلالة الإحصائية، بينما كان البحث في علم النفس قبل العام ١٩٤٠ يعتمد على الأسلوب غير الكمي. وقد بدأ استخدام قيم r بدلاً من القيم الحرجة خلال الفترة ١٩٤٠ –١٩٥٥. وقد كانت متوسطات عدد المقالات الإمبريقية التي استعملت اختبارات الدلالة الإحصائية من العام ١٩١١ وحتى عام ٢٠٠٠ هي: ١٧٪ للأعوام ١٩١١ –١٩٢٩ ، ووصلت إلى ٨٥٪ في عام ١٩٦٠، ثمّ تجاوزت ٩٠٪ في الأعوام من ١٩٧٠ وما بعد. وقد أوضحت هذه الدراسة أنّ اختبار الدلالة الإحصائية حالياً هي اللغة الشائعة للإحصاء الاستدلالي في جميع مجلات الجمعية النفسية الأمريكية APA. كما توصلت هذه الدراسة إلى أنّ الإعادة مع البحث الموسع، وارتباطه بتحليل ميتا، هي وسيلة أفضل من اختبارات الدلالة الإحصائية لتقدّم ونمو المعرفة التراكمية في علم النفس.

(٢) "مسح وطني من رابطة البحث التربوي الأمريكية عن إدراك الأعضاء لطبيعة ومعنى الدلالة الإحصائية"

قام الباحث ميتاق Mittag(1999) بدراسة للكشف عن إدراك أعضاء رابطة البحث التربوي لمعنى وقيمة الدلالة الإحصائية. وقد شملت عينة الباحث عينة عشوائية من ٤٪ من أعضاء جمعية البحث التربوي الأمريكية AERA من (١٢) قسماً مختلفاً من العدد الكلي للأعضاء وهو ١١٢٧ عضوا، وقد كانت الأداة عبارة عن استطلاع من ورقة ذات وجهين، حيث سأل المسح عن الجنس، وما إذا كان المشارك حاصلاً على الدكتوراة، وطبيعة العمل الأساسية، احتوى المسح على (٢٩) فقرة، وكانت الإجابات من خمس فئات استجابة، وقد ظهر من توصيف عينة الدراسة أنّ ٨٣,٦٪ منهم يحملون درجة الدكتوراة، وأنّ ٦٥,٨٪ منخرطون في العمل الأكاديمي في الجامعات.

وافق المستجيبون للمسح على أنّه لا بد من أن تستخدم كلمة (دال إحصائياً) بدلاً من كلمة (دال)، ووافقوا أيضاً على أنّ هذا المفهوم (دال) سيستمر وقتاً طويلا، لكنّهم لم يوافقوا على أنّ الدلالة الإحصائية يجب أن تمنع، كما أظهرت الدراسة أنّ المستجيبين كانوا محايدين بخصوص تأثير حجوم العينات في اختبارات الدلالة الإحصائية، وأنّ أي فرضية صفرية يمكن أن ترفض عند حد ما من حجم العينة. كما أظهرت الدراسة أنّ المستجيبين لم يميلوا إلى أنّ قيم (r) تقيس بشكل مباشر حجم تأثير الدراسة أو أنّ الإخفاق في الحصول على نتيجة دالة إحصائية تعني أنّ النتائج لم تكن مفيدة أو مهمة، وقد رأى الباحث أنّ مثل هذه النتيجة مخيبة للآمال. كما أظهرت الدراسة أنّ المستجيبين وافقوا على أنّ القيم غير الدالة إحصائياً من الممكن أن تكون مهمة للغاية، ويرى الباحث أنّ هذا الشيء مثير للفرح والسرور، وعندما سئل المستجيبون ما إذا كانت قيم (r) تقوّم تكرار النتيجة، وجدت الدراسة أنّ الاستجابات كانت محايدة، وهذا في رأي الباحث في تلك الدراسة محبط.

وقد توصلت الدراسة للنتائج التالية:

١ - الباحثون لم يدركوا بأنّه لأي نتائج عينة، فإنّ الفرضية الصفرية دائماً ترفض عند حد معيّن من حجم العينة.

٢ - الباحثون لا يفسرون بشكل صحيح النتائج غير الدالة إحصائياً.

٣ - لم يحسم الباحثون أنّ اختبارات الدلالة الإحصائية لا تفسّر إمكانية تكرار النتيجة في دراسات مستقبلية.

٤ - أنّ هذه النتيجة تمثّل لبّ الانتقاد الحالي للممارسة البحثية.

(٣) "تحسين وتوضيح نتائج الأبحاث، وفائدة مؤشرات حجم التأثير كاختبارات مكملة لاختبارات الدلالة الإحصائية".

قام الباحث ثومبسون (Thompson(1999a بفحص مدى إشارة مجلة (الأطفال الاستثنائيون) (Exceptional Children) لحجم التأثير. شملت عينة الدراسة المقالات المنشورة من المجلد رقم (٦٣) في خريف عام ١٩٩٦ إلى مجلد (٦٤) في ربيع عام ١٩٩٨م، حيث كان مجموع المقالات المنشورة (٢٣) مقالاً. وجد الباحث أنّ (٢٠) مقالاً من الـ (٢٣) مقال لم تذكر أي مؤشر لحجم التأثير. وتوصل الباحث إلى ضرورة تضمين الدراسات لمؤشرات عن حجم التأثير وفقاً لتعليمات الجمعية النفسية الأمريكية APA.

(٤) "استخدام اختبارات الدلالة الإحصائية واختبارات تحليلية أخرى في مجلة كلية علم النفس: مراجعة الممارسات واقتراح بدائل".

قام الباحثان سنايدر وثومسون (Snyder and Thompson (1998 بدراسة سبع مجلدات لمجلة (كلية علم النفس الربع سنوية) (School Psychology Quarterly). وقد ركزت الدراسة على تفسير نتائج الدراسات خاصة في استعمال اللغة غير المناسبة. وقد وجد الباحثان أنّ المؤلفين المعاصرين مستمرون في استخدام وتفسير اختبارات الدلالة الإحصائية بشكل غير مناسب، وأوصيا بضرورة الاهتمام باستعمال اللغة الدقيقة مثل كلمة (دال إحصائياً) بدلاً من كلمة (دال) فقط، كما توصلت الدراسة إلى ضرورة تضمين مؤشرات إحصائية أخرى عند تفسير النتائج.

(٥) "استخدام اختبارات الدلالة الإحصائية وتحليل الثبات في أبحاث الإرشاد المنشورة".

هذه الدراسة للباحثين ثومبسون وسنايدر (Thompson and Snyder(1997a عبارة عن دراسة ذاتية نقدية لـ (مجلة النمو والإرشاد) (Journal of Counseling and Development). هدفت الدراسة إلى القيام بدراسة المجلة التي صاغت رسالتها على أنّها (جزء من ضمير المحترفين)، وأنّها تحاول أن تقدّم نفسها كـ (سجل عالمي للاستشارات المحترفة) .

هدف الباحثان من دراستهما بحث قضيتين منهجيتين ومعاصرتين هما:

- استخدام وتفسير اختبارات الدلالة الإحصائية.

- معنى وطريقة تقويم ثبات الدرجات للمقاييس المستعملة في التساؤلات الكمية والجوهرية.
- ثمّ وصف الممارسات المتعلقة بتلك القضيتين في مقالات مجلة (النمو والإرشاد).

حدّد الباحثان عينة الدراسة بجميع الدراسات الكمية المنشورة في مجلة (النمو والإرشاد) عام ١٩٩٦، حيث شملت العينة خمساً وعشرين دراسة كمية نشرت في المجلة. وقد قام الباحثان في كل مقال من مقالات الدراسة بحساب التباين المفسّر لحجم التأثير r2، وقد أنتج ذلك الإجراء ما معدله ١٠,٩ حجم تأثير لكل مقال (SD=13.4) (مجموع التأثيرات الإجمالية هي ٢٧٤). علماً بأنّ الباحثين لم يتمكنا من حساب حجوم التأثير في بعض المقالات بسبب عدم توفر المعلومات اللازمة، حيث أنّ الباحثين الأصليين يكتفون في بعض الأحوال بالإشارة إلى أنّ النتائج لم تكن دالة إحصائياً دون التزويد بأي معلومات إضافية. كما أنّهما قاما بفحص جميع المقالات لتحليل ثبات الدرجات، حيث صنّف الباحثان وسائل تقرير ثبات الدرجات إلى ثلاثة أقسام:

(أ) عدم وصف أي معاملات ثبات للقياسات المستعملة.

(ب)عرض معاملات ثبات من دراسات سابقة.

(ج) حساب معاملات الثبات الفعلية للدراسة.

وقد وجد الباحثان أيضاً في (١٥) دراسة من الـ (٢٥) دراسة أنّ الباحثين أشاروا إلى حجم تأثير واحد على الأقل (مثل Cohen's d , Omega2 , eta2 , r2) ، وفي دراستين فقط من الـ (٢٥) دراسة ضمّن المؤلفون حجوم التأثير مع تفسيرها، كما وجدا أنّ متوسط التباين المفسّر للتأثيرات الـ (٢٧٤) المتضمنة في الـ (٢٥) مقالاً هي ٠,١٤٨ ، كما وجدا أنّ العديد من الباحثين أشاروا إلى النتائج بأنّها (دالة) للإشارة إلى أنّها (دالة إحصائياً).

أمّا فيما يخص تقويم ثبات الدرجات، فقد وجد الباحثان بأنّ دراسة واحدة من الـ (٢٥) مقالاً عرضت معامل ألفا للبيانات الفعلية المنشورة في المقال، و(٨) مقالات أشارت إلى معاملات الثبات المحسوبة فعلياً إضافة للمعاملات من دراسات سابقة، وفي (١٣) مقالاً أشار الباحثون إلى معاملات الثبات من دراسات سابقة فقط، وذلك بسبب الاعتقاد أنّ الثبات هو للمقياس وليس للدرجات نفسها.

ولاحظ الباحثان أنّ كثيراً من الباحثين يصرون على مساواة عدم احتمالية النتيجة بقيمتها، وأنّ كثيراً من الباحثين يصرون على الاعتقاد بأنّ الدلالة الإحصائية تقوّم ثبات النتيجة، بالإضافة إلى أنّ كثيراً من الباحثين يصرون أيضاً على الاعتقاد بأنّ عدم احتمالية النتيجة تساوي مقدار تأثيرات الدراسة. أمّا فيما يخص تقويم الثبات في الدراسات السابقة، فقد وجد الباحثان بأنّ (١٣) دراسة في مجلة (النمو والإرشاد) اعتمدت في ثبات الدرجات على

دراسات سابقة، وأنّ (٨) دراسات فقط اعتمدت على ثبات الدرجات التي بين يديها إضافة إلى ثبات دراسات سابقة.

وقد توصل الباحثان إلى ضرورة أن يبدأ الباحثون في عرض وتفسير حجوم التأثير لتساعدهم في تفسير النتائج، وكذلك في التفسيرات التي يقوم بها قرّاء الأبحاث. كما أنّهما توصلا إلى أنّ المؤلفين يحتاجون إلى التمييز بأنّ الثبات يخص مجموعة من الدرجات وليس الاختبار نفسه.

٦) "ممارسة اختبارات الدلالة الإحصائية في مجلة التربية التجريبية"

قام ثومسون وسنايدر (Thompson and Snyder(1997b بفحص ثلاث من الوجهات الموصى بها في جميع الأبحاث الكمية في المجلدين (٦٣) و (٦٤) للعامين الأكاديميين (١٩٩٤/١٩٩٥ و ١٩٩٥/١٩٩٦) بمجلة التربية التجريبية (The Journal of Experimental Education). وقد شملت عينة الدراسة (٢٢) مقالاً منشوراً استخدمت اختبارات الدلالة الإحصائية. وقد قام الباحثان بحساب حجم التأثير لكل الدراسات وذلك للتحقق من افتراضهما بأنّ الباحثين يعتمدون على الاختبارات الإحصائية دون التحقق من أنّ تلك النتائج ربما لا تكون مفيدة أو جديرة بالاهتمام، حيث أنّهم يختبرون الدلالة الإحصائية فقط دون حساب حجوم التأثير. وقد افترض الباحثان أيضاً أنّ بعض الباحثين يحصلون على نتائج دالة إحصائياً بسبب حجم العينة الكبير، بينما نتائجهم تستلزم حجوم تأثيرات صغيرة فقط.

وقد وجد الباحثان أنّ بعض الباحثين لمقالات عينة الدراسة (٢٢ مقالاً) أشاروا إلى عدد فوق العادة لاختبارات الفرضيات لبيانات عينة واحدة، مع اهتمام محدود بشكل واضح للتأثير الناتج عن الخطأ من النوع (I). كما وجد الباحثان بأنّ اللغة المستخدمة في كل المقالات الـ (٢٢) تضمنت وصف النتائج بأنّها (دالة) وليست (دالة إحصائياً)، كما أنّ بعضها تضمنت لغة مثل (تقترب من الدلالة) ، (كانت تقريباً دالة)، (لا تدنو من الدلالة)، (تأثير المعالجة اقتربت من مستوى a المحدّد).

كما لم يجد الباحثان ولا دراسة واحدة من المقالات الـ (٢٢) تضمنت نتائج لتحليل التكرار الداخلي، علماً بأنّ اختبارات الدلالة الإحصائية لا تقوّم تكرار النتائج، ممّا يعني أنّ تحاليل أخرى مطلوبة لعرض هذه القضية المهمة. كما وجد الباحثان أنّ (٨) مقالات من المقالات الـ (٢٢) لم تتضمن الإشارة إلى حجوم التأثير بشكل مصاحب لاختبارات الدلالة الإحصائية، وأنّ (٦) مقالات تضمنت الإشارة إلى إحصاءات مرتبطة مباشرة بالتباين المفسّر مثل (r2 , r)، ومع ذلك لم يكن واضحاً بأنّ أولئك الباحثين قد فسّروا تلك النتائج الارتباطية

كحجوم تأثير لغرض الحكم على مقدار التأثيرات التي أشاروا إليها. ووجد الباحثان أنّ (٤) مقالات تضمنت حجم تأثير واحد على الأقل، مع أنّ نفس هذه المقالات تتضمن اختبارات دلالة إحصائية أخرى لم تعرض لها أي تقديرات لحجوم التأثير. كما وجد الباحثان أنّ (٤) مقالات كانت مقاييس حجوم التأثير محور عرض النتائج وتفسيرها. وقد كانت تلك المقاييس هي: r , Cohen's d , eta2 . كما وجدت الدراسة أنّ حجوم العينات في الأبحاث والمقالات التي درستها تتراوح من (٢٦) إلى (١٣١٢١) حالة، بمتوسط (م=١١٠٤، وانحراف معياري= ٣,١٩٦١). وقد كانت ثلاث مقالات أحجام عيناتها فوق ١٠٠٠ حالة.

(٧) "الاعتقادات الخاطئة حول حجم العينة، والدلالة الإحصائية، و تأثير المعالجة".

قام الباحثان ويلكرسون وأولسون Wilkerson and Olson (1997) بتقديم أداة مكونة من (٦) أسئلة مصممة لقياس تفسير العلاقات بين حجم العينة وتأثير المعالجة، وحجم العينة والخطأ من النوع (I) والخطأ من النوع (II)، وقد قدّمت الأداة لـ (٥٢) طالب دراسات عليا في جامعة Idaho ، حيث كان (٢٠) منهم يحضرون درجة الدكتوراة في الفلسفة Ph.D. (٣٩٪)، و (١٤) يحضرون درجة الدكتوراة في التربية Ed. D. (٢٧٪)، و (١٦) يحضرون درجة الماجستير (٣١٪)، بينما (٢)منهم لم يحدّدا استجابتهما في خانة الدرجة التي يدرسانها (٤٪).

كان السؤال الرئيس في الدراسة هو : كيف يفسر الباحثون العلاقة بين حجم العينة وتأثير المعالجة، إذا كانت قيم النتائج الدالة إحصائياً متساوية؟ وكيف يفسر الباحثون العلاقة بين حجم العينة والخطأ من النوع (I)؟ وكيف يفسر الباحثون العلاقة بين حجم العينة والخطأ من النوع (II). تم توزيع الاستبانة للطلاب المشاركين في ثلاث فصول خلال الفصل الصيفي لعام ١٩٩٦م، حيث أعطوا (٥) دقائق من وقت الحصة لإكمال وإرجاع الاستبانات. الأداة كانت عبارة عن صفحة واحدة.

وقد وجد الباحثان في هذه الدراسة أنّ طالباً واحداً فقط من الـ (٥٢) أي بنسبة (١,٩٪)كان مدركاً بأنّه بإعطاء دراستين مختلفتين ومتساويتين في قيمة (r)، فإنّ الدراسة التي بعدد أقل من (n) تزودنا ببرهان أفضل عن تأثير المعالجة. كما وجدا أنّ (٦) طلاب من الـ (٥٢) أي بنسبة (١١,٥٪) ميّزوا أنّ احتمالية الخطأ من النوع(I) لا يعتمد على حجم العينة. وكما وجدا أنّ (٢٥) طالباً من الـ (٥٢) أي بنسبة (٤٨٪) يفهمون أنّ احتمالية الخطأ من النوع (II) تقل مع زيادة حجم العينة.

توصل الباحثان إلى أنّه بغض النظر عن الانتشار الواسع لاختبارات الدلالة الإحصائية، فإنّ هناك سوء فهم مرتبط بتفسيرات اختبارات الدلالة الإحصائية، وأنّ نتائج

دراستهم تؤكد وجود سوء فهم للعلاقة بين حجم العينة والدلالة الإحصائية، وأنّ المستجيبين وضعوا ثقة أكبر في نتائج الدراسات التي أحجام عيناتها كبيرة مقارنة بالدراسات التي أحجام عيناتها صغيرة، بغض النظر عن المعيار الذي تعتمد عليه تلك الثقة. ويرى الباحثان بأنّ دراستهما تزود بمعلومات عن مستوى فهم الباحثين في الدراسات العليا للقضايا الأساسية في تفسير نتائج الأبحاث. ويرى الباحثان بأنّ على الرغم من أنّ هذه الدراسة محدودة، إلاّ أنّهما يعتقدان بأنّ هذا الخطأ في التفسير منتشر في العلوم الاجتماعية، وقد ذهب الباحثان إلى أنّ المزيد من فهم العلاقة بين حجم العينة والدلالة الإحصائية سوف يقود إلى تفسيرات أكثر دقة في نتائج البحث، وفي وضع المعرفة في العلوم الاجتماعية التي سوف تستفيد من هذا التطوّر والتحسّن في التفسيرات.

٨) "كشف أسطورة نتيجة عالية الدلالة: حجم التأثير في الأبحاث التربوية".

قام الباحث بلوكير (Plucker(1997 بدراسة ثلاث مجلات علمية ربع سنوية وأربعين مقالاً في مجلات أخرى نشرت خلال الخمس سنوات الأخيرة من البحث. وقد وجد الباحث أنّ تلك الأبحاث لا تتضمن في العادة أي مقاييس لحجوم التأثير، كما وجد أنّ عبارة (عالية الدلالة) موجودة في أغلب تلك المجلاّت والمقالات المنشورة. وتوصل الباحث إلى أهمية إدراك الباحثين لمفهوم الدلالة الإحصائية وذلك لتحسين تفسير النتائج.

٩) "الدلالة العملية: مفهوم حان وقته".

قام الباحث كيرك (Kirk(1996 بدراسة استخدامات اختبارات دلالات الفروض الصفرية في أربع دوريات علمية للجمعية النفسية الأمريكية والمنشورة عام ١٩٩٥م. وجدت الدراسة أنّ ٧٧٪ من المقالات تحتوي على مقياس واحد أو أكثر من مقاييس حجم التأثير. كما وجدت الدراسة أنّ أكبر عدد من مقاييس حجم التأثير المستخدمة هي التباين المفسّر، علماً بأنّ الباحث لم يتمكن من تحديد أنواع تلك المقاييس (مثل Eta2 , R2 ... وغيرها). ومع ذلك فإنّ الباحث عندما ضمّ مقياس (R2) مع بقية مقاييس التباين المفسّر مثّلت نسبة المجموع ٦٠٪ من مؤشرات حجم التأثير المستخدمة في الأبحاث. وقد ذهب الباحث إلى أنّ ارتفاع نسبة الإشارة إلى حجم التأثير في الدراسات والتي بلغت ٧٧٪ قد لا يعود بالضرورة إلى إدراك الباحثين لأهمية تلك المؤشرات، ولكن بسبب أنّ برامج الحاسب الآلي عادة ما تعطي بعض تلك المؤشرات في نتائجها مثل ما يحدث تلقائياً من عرض لقيمة R2 مع جميع نتائج الانحدار، وذهب الباحث إلى أنّ هذا هو سبب ارتفاع نسبة الإشارة إلى مؤشرR2 . وقد توصل الباحث أيضاً إلى ضرورة حساب وعرض قيم الدلالة العملية.

(١٠) "حجم التأثير كبديل لاختبار الدلالة الإحصائية".

قام الباحث ماكلين McClain(1995) بمراجعة مجلة علم النفس الإكلينيكي (Journal of Clinical Psychology) حيث كانت عينة الدراسة (١١٣) مقالاً نشرت بالمجلة عام ١٩٩٣م. كانت نسبة المقالات التي احتوت على اختبار F أو t هي ٤٨٪. وقد حصل الباحث من تلك العينة على ١٢٨ إحصاءة اختبار. وقد استعمل الباحث عدد (٣٣) من قيم إحصاءة (t) و (٤٨) من قيم إحصاءة F ذي الاتجاه الأحادي، وذلك لحساب حجم التأثير. قام الباحث بالحصول على : إحصاءة الاختبار، التحليل الإحصائي، حجم العينة، حجم التأثير المتضمن في المقال، قيم (r) المحسوبة، درجات الحرية وذلك من (٥٤) مقالاً استخدمت إمّا اختبار (F) أو اختبار (t)، حيث قام الباحث بحساب قيمة مربع أوميغا المصححة و مربع إيتا غير المصححة كحجوم تأثير. وقد استبعد الباحث بعض إحصاءات الاختبارات بسبب فقد معلومات مثل حجم العينة، أو درجات الحرية. وقد استخدم الباحث رسم الانتشار لتوضيح العلاقة بين حجوم التأثير المصححة وغير المصححة مع حجم العينة.

وجد الباحث عدّة مشاهدات مهمة منها: أنّ من (١٢) إحصاءة اختبار مفحوصة كانت (٩) إحصاءات (٧٪) هي التي صوحبت بوصف لحجوم التأثير. وبشكل عام فقد أشير إلى حجوم التأثير في (٥) مقالات من (٥٤) مقالاً أي بما يعادل (٩٪). كما وجد الباحث أنّ من الـ (٥٤) مقالاً المراجعة، استخدم (٤٤) مقالاً أي ما نسبتها (٨١٪) عبارة (دال) و (دال إحصائياً)، و (٣) مقالات أي ما نسبتها (٦٪) استخدمت عبارة (دال إحصائياً) في وصف الدلالة الإحصائية، كما توصلت الدراسة إلى عدم وجود ارتباط بين حجوم التأثير و أحجام العينات.

وتوصل الباحث إلى أنّ الباحثين لا يزالون مستمرين في الاعتماد الشديد على اختبار الدلالة الإحصائية في تقويم نتائج البحث. كما أنّ استخدام الغالبية (٨١٪) للغة (دال) للإشارة إلى نتائج الدلالة الإحصائية يستلزم أنّ تلك النتائج ذات معنى. كما استخلص الباحث من خلال معاملات الارتباط التي بلغت ($r = -0.1007$) و ($r = -0.1189$) بين حجم العينة وحجم التأثير المصحح أو غير المصحح، أنّ حجوم التأثير تعكس صورة أوضح للعلاقة الموجودة في البيانات لأنّها لا تُحجب بواسطة حجم العينة مثل الإحصاءات الناتجة من اختبارات الدلالة الإحصائية. وتوصل الباحث إلى أنّ فهم واستخدام حجم التأثير سوف يمنع سوء تفسير الاختبارات الإحصائية كمؤشر للأهمية، وأنّ البحث النفسي مثل الذي تمت مراجعته في دراسة هذا الباحث، يمكن أن يفسّر بطريقة أفضل باستخدام حجوم العينة.

(١١) "التقويم غير الملائم للدلالة الإحصائية".

قام الباحث هاركوم Harcum(1989) بمراجعة (١٥) مقالاً من مجلة APA ، حيث وجد أنّ (٩) مقالات وجدت فيها دلالة إحصائية. وعندما درس الباحث هذه المقالات الـ (٩)، وجد أنّ عبارات مثل (عالي)، (بسهولة)، (بوضوح)، (جداً) استخدمت عندما كانت قيمة (a) أقل من ٠,٠١، وأنّ عبارات مثل (هامشية)، (تقترب من الدلالة) استعملت عندما كانت (a) بين ٠,١ و ٠,٠٥. كما وجد أنّ الباحثين يستعملون (كبير)، (عظيم)، (قوي) عند الإشارة إلى قيم (r) المنخفضة (r <0.001).

توصل الباحث إلى أنّ نتائج الدراسة تدعم البراهين بأنّ المشكلة موجودة، وهي عرض قيم الدلالة الإحصائية لوحدها دون الإشارة إلى مقدار أو أهمية العلاقة الموجودة. هذا يعني أنّ استخدام مستويات الاحتمالية لاختبارات الدلالة الإحصائية كمقاييس لقوة العلاقات خطأ شائع جداً.

(١٢) "تحيّز المحكمين للنتائج الدالة إحصائياً: إعادة اختبار".

قام الباحثان فاقلي و ماك كنّي Fagley and McKinny(1983) بإعادة اختبار مقالات أتكينسون، فورلونغ ووامبولد(١٩٨٢) وتوصلا إلى أنّ المحكمين متحيزين تجاه نتائج الدلالة الإحصائية، وأنّ تحليل القوة الإحصائي يوضح أنّ القوة منخفضة لدراساتهم، وأنّ انخفاض القوة في الدراسات التي لا تشير إلى دلالة إحصائية سبب كافٍ للتوصية بعدم النشر.

ثانياً - الدراسات غير التطبيقية

(١) "مراجعة لآخر الأدبيات فيما إذا كانت اختبارات الدلالة الإحصائية يجب أن تدان"

قام الباحث بوردينيسكي (Burdenski(1999 بدراسة ركزت على مراجعة الأدبيات التي تدعم أو تعارض استخدامات اختبارات الدلالة الإحصائية ولخصت الدراسة ثلاثة قصور للدلالة الإحصائية. الأول أنّ قيم (r) بنفسها لا يمكن استخدامها كمؤشرات لحجم التأثير، الثاني أنّ النتائج غير المحتملة لا تعني الأهمية، الثالث أنّ قيم (r) غير مسئولة عن القضية المهمة لثبات النتائج لأنّ اختبارات الدلالة الإحصائية لا تختبر إمكانية أن تكون نتائج العينة تحدث في المجتمع الإحصائي. كما تضمنت الورقة مناقشة ما يمكن أن تقوم به اختبارات الدلالة الإحصائية للفروض الصفرية وما لا يمكن أن تقوم به.

(٢) "ما بعد اختبارات الدلالة الإحصائية: أهمية استخدام تقديرات أخرى لتأثيرات المعالجة لتفسير تقويم النتائج".

قام الباحث كيلو (Kellow(1998 بدراسة تقديرات تأثيرات المعالجة الضرورية لتفسير الاختلافات بين المجموعات، وقد توصلت الدراسة إلى أنّ العديد من طلاب القياس لا يزالون يفكرون في استخدام اختبارات الدلالة الإحصائية دون الانتباه لقصورها وحدودها، وقد توصل الباحث إلى أنّ كثيراً من البرامج لا تزال مستمرة اعتماداً على حجم العينة المستخدمة لتقويم البرنامج وليس بناء على الفعالية الحقيقية للبرنامج، وأنّ ذلك يحدث بسبب أن المقومين ليسوا مهتمين ولا مدركين بأنّ اختبارات الدلالة الإحصائية والتي يعتمدون عليها في مثل تلك التقويمات والتوصيات هي من صنيع حجم العينة الكبير، بينما تكون الفروق الفعلية لحجم العينة ليس لها معنى في الواقع الفعلي. وقد عرض الباحث نماذج توضح أنّ مقدار التباين المحسوب لـ ($r2$) يمكن أن يكون ثابتاً ، ومع ذلك يعطي نتائج ذات دلالة إحصائية مختلفة بسبب الفرق البسيط في حجم العينة. وقد توصل الباحث إلى تحديد مفهوم الدلالة الإحصائية وأنّها عبارة عن تقويم لاحتمالية الحصول على البيانات إذا كانت الفرضية الصفرية صحيحة. كما أوصى كيلو جميع الباحثين والمقومين خاصة المبتدئين منهم، بأنّهم حتى لو شعروا بأنّهم تائهون في بعض تفصيلات حجوم التأثير وأي حلول أخرى، فإنّ عليهم ألاّ يدعوا استخدام بعض تلك الطرق لتقدير الدلالة العملية وحجم التأثير. وأوصى أيضاً بأنّه حتى لو شعر الباحث بأنّه لا يعرف إلاّ طريقة واحدة فقط من طرق تقدير حجم التأثير فإنّ عليه أن يستعملها ، المهم أن يقلّل اعتماده على اختبارات الدلالة الإحصائية بمفردها فقط.

(٣) "ماذا لو لم يكن هناك مزيد من التخاصم حول اختبارات الدلالة الإحصائية؟"

قام بالدراسة ليفين Levin(1998) ، حيث تضمنت الدراسة نقاطاً موجهة للمدافعين ضد استبدال اختبارات الفرضيات الإحصائية باستراتيجيات جديدة لتحليل البيانات. توصلت الدراسة إلى أنّ الاقتراحات التي عادة ما يوصى بها كبدائل ليست تامة، وأنّ تطويرات متعدّدة على اختبارات الفرضيات يمكن أن تطبّق لجعل العملية واستنتاجاتها جديرة بالثقة.

(٤) "خمسة أخطاء منهجية في البحث التربوي"

قام ثومبسون Thompson(1998c) بدراسة خمسة أخطاء منهجية تحدث في البحث التربوي، وتوصلت الدراسة إلى أنّ أحد هذه الأخطاء تخص الدلالة الإحصائية وهي التفسير الخاطئ للدلالة الإحصائية وما يتعلق بها من فشل تقرير وتفسير حجم الأثر الممثّل في جميع التحاليل الكمية.

(٥) "اختبار الدلالة الإحصائية: من الروتين إلى الطقوس".

قام الباحث كيستر Keaster(1998) بدراسة الأدبيات حول انتقاد الممارسة الحالية للباحثين والناشرين في سياق اختبارات الدلالة. وتوصل إلى أنّ المشكلة المعتبرة عندما يحاول الباحث لعمل أكثر من مجرد تقرير وجود علاقة ملاحظة، وأنّ أكثر الباحثين يفترضون أنّ الفروق وحجم تلك الفروق تبرهن أو على الأقل تؤكد فرضيات البحث، كما اقترحت الدراسة بأنّ على المؤلفين و الباحثين عدم محاولة تفسير قيم الدلالة الموجودة، وأنّ هناك حاجة إلى عودة اختبارات الدلالة إلى مكانها الصحيح في العملية العلمية.

(٦) "دور اختبارات الدلالة الإحصائية في البحث التربوي".

قام الباحثان ماكلين و إيرنست McLean and Ernest(1998b) بمناقشة كيف أنّ اختبارات الدلالة الإحصائية كأساس وحيد لتفسير النتائج تعتبر ممارسة زائفة، بينما يمكن أن تكون اختبارات الدلالة مفيدة كواحدة من عدّة معايير يجب أن تؤسس في الوضع الإمبريقي. وتوصلت الدراسة إلى أنّ اختبارات الدلالة الإحصائية تزود ببراهين بأنّ الأحداث لا تحدث بالصدفة، لكنّها لا تزود ببراهين عن معنى الأحداث.

(٧) "هل اختبار الدلالة الإحصائية ما يزال محتفظاً بفائدته؟"

قام بالدراسة الباحثان ماكلين وإيرنست (McLean and Ernest(1997، وهدفت دراستهما إلى الدفاع عن مكانة الدلالة الإحصائية وإثبات أنّها لم تفقد فائدتها. وقد توصلت الدراسة إلى أنّ أدبيات طرق البحث في السنوات الحالية تضمنت هجوماً كاملاً على اختبارات الدلالة الإحصائية، واعترفت الدراسة بأنّ اختبارات الدلالة الإحصائية يجب أن تؤخذ في الاعتبار بجمعها مع معايير أخرى. كما توصلت الدراسة إلى أنّ الدلالة الإحصائية تزود ببراهين عن الأحداث التي لا تحدث بالصدفة، ولكنّها لا تزود بمعلومات حول معنى (الدلالة العملية) للحدث، ولا ما إذا كان الحدث ثابتاً. ولذا فإنّ الدلالة الإحصائية يجب أن تكون مرافقة لاختبارات الدلالة العملية للحكم على الأحداث وثباتها. وذهبت الدراسة إلى أنّ من أسباب فقد الدلالة الإحصائية أهميتها كبر حجم العينات، ولذا فإنّ اختبارات الدلالة الإحصائية سوف تكون المعيار الوحيد المفيد للباحثين الاستدلاليين عند اختبار العدد القليل من البيانات، ومع ذلك فقد أوصت الدراسة الباحثين بعدم إهمال الأساليب الأخرى مثل حدود الثقة.

(٨) "اختبارات الدلالة الإحصائية يجب أن تتوقف في البحث التربوي".

قام الباحث مينون (Menon(1993 بدراسة بعض الأساطير الشائعة حول اختبارات الدلالة الإحصائية ، والأخطاء الممكنة والمساهمات الضارة لاختبارات الدلالة الإحصائية في البحث التربوي. توصلت الدراسة إلى أهمية التوقف عن استخدام اختبارات الدلالة الإحصائية.

(٩) "تقويم الدلالة الإحصائية باستخدام التقديرات الصحيحة وغير الصحيحة لحجم التأثير".

قام بالدراسة الباحثان سنايدر ولاوسون (Snyder and Lawson(1992، حيث هدفت الدراسة إلى تناول مقاييس حجم التأثير. كما تناولت الدراسة أسباب تشجيع المختصين في البحوث التربوية استخدام حجم التأثير كمساعد تفسيري، والأنواع المختلفة من قياسات حجم التأثير، مع تصنيفها إلى مجموعات حسب: التحيّز وعدم التحيّز ، مؤشرات المجتمعات والعينات ، ومؤشرات نماذج التصاميم الثابتة مقابل العشوائية، وتوصلت الدراسة إلى أنّ مقاييس حجم التأثير إذا استخدمت بطريقة صحيحة فإنّها تعتبر مساعدة للباحثين الذين لا يرغبون في الاعتماد فقط على اختبارات الدلالة الإحصائية في تفسير النتائج المهمة.

ج) التعليق على الدراسات السابقة

من خلال استعراض الدراسات السابقة يمكن ملاحظة ما يأتي:

١ - تنوعت الدراسات السابقة بين دراسات تطبيقية حاولت اختبار فرضيات محدّدة مثل دراسات: هوبارد وريان Hubbard and Ryan(2000)، وميتاق Mittag(1999)، وثومبسون Thompson(1999a)، وسنايدر وثومبسون Snyder and Thompson (1998) وثومبسون وسنايدر Thompson and Snyder(1998)، وثومبسون وسنايدر Thompson and Snyder(1997a)، وويلكرسون وأولسون Wilkerson and Olson (1997)، وبلوكير Plucker(1997)، وكيرك Kirk(1996)، والصائغ (١٤١٥)، وماكلين Mclain(1995)، ونور (١٤١٣)، والنجار (١٤١١)، وهاركوم Harcum(1989)، والصياد (١٩٨٨)، وفاقلي وماك كنّي Fagley and McKinny(1983)، ودراسات أخرى أساسية نظرية مثل دراسات: بوردينسكي Burdenski(1999) ، وكيلو Kellow(1998)، وليفين Levin(1998) ، وثومبسون Thompso(1998)، وكيستر Keaster(1998)، وماكلين وإرنست Mclean & Ernest(1998)، وماكلين وإرنست McLean & Ernest (1997) ، ومينون Menon(1993)، وسنايدر ولاوسون Snyder & Lawson(1992)، والصياد (١٩٨٩) وهذا التنوع بمجمله يعطي صورة أكثر تكاملاً عن مشكلة الدراسة.

٢ - أفادت دراسات نور(١٤١٣) والنجار(١٤١١) والصائغ(١٤١٥) بأنّ قيمة مؤشرات الدلالة العملية لاختبارات الدلالة الإحصائية كانت منخفضة في رسائل الماجستير بكلية التربية بجامعة أم القرى، وقد أفادت دراسة الصياد (١٩٨٨) والتي شملت عينات من إحدى الجامعات المصرية وكذلك بعض الدوريات العربية بنفس النتيجة التي تتضمن انخفاض قيم مؤشرات الدلالة العملية لاختبارات الدلالة الإحصائية ممّا يعني أنّ انخفاض قيمة مؤشرات الدلالة العملية موجودة في البحث التربوي العربي وليس فقط المحلي.

٣ - كما أفادت دراسة نور(١٤١٣) والصياد(١٩٨٨) بأنّ حجم العينة في البحث التربوي العربي كبير جداً ومتضخم.

٤ - بعض هذه الدراسات أثارت مشكلات أكثر من تقديم حلول، مثل دراسات: بوردينسكي Burdenski(1999)، وثومبسون Thompson(1999)، وسنايدر وثومبسون Snyder & Thompson(1998)، وبلوكير Plucker(1997)، وثومبسون وسنايدر Thompson & Snyder(1997)، ومينون Menon(1993)، وفاقلي و ماك كنّي Fagley and McKinny(1983).

٥ - بعض الدراسات تناولت مشكلات الدلالة الإحصائية بشكل غير مستقل ولا مركز، بل تناولتها كمشكلة ضمن عدّة مشكلات عالجتها الدراسة، مثل دراسات كل من: ثومبسون Thompson(1999)، وثومبسون وسنايدر Thompson & Snyder(1998)، وثومبسون Thompson(1998)، وثومبسون وسنايدر Thompson & Snyder(1997).

٦ - بعض الدراسات السابقة قدّمت حلولاً جزئية وذلك حسب أهداف الدراسة المحدّدة مثل دراسات كل من: ثومبسون Thompson(1999)، وسنايدر وثومبسون Snyder & Thompson (1998)، وكيلو Kellow(1998)، وماكلين وإرنست McLean & Ernest (1997)، وكيرك Kirk(1996)، وماكلين McLean(1996)، وسنايدر ولاوسون Snyder & Lawson(1992).

٧ - أفادت الدراسات السابقة بأنّ الجدل في قضية الدلالة الإحصائية لا يزال قائماً بين مؤيد ومعارض، مثل دراسات: بوردينسكي Burdenski(1999)، وليفين Levin(1998)، وماكلين وإرنست McLean & Ernest (1997).

٨ - أفادت بعض الدراسات السابقة مثل: ثومبسون Thompson(1999)، وسنايدر وثومبسون Snyder & Thompson (1998)، وليفين Levin(1998)، وثومبسون وسنايدر Thompson & Snyder(1997) بأنّ الاقتراحات التي طرحت سابقاً لم تؤد إلى أثر ملموس، كما أنّها إلى حد كبير غير تامة.

٩ - الدراسات السابقة مثل: ميتاق Mittag(1999)، وبوردينسكي Burdenski(1999)، وكيستر Keaster(1998)، وكيلو Kellow(1998)، وثومبسون وسنايدر Thompson & Snyder(1997)، وويلكرسون وأولسون Wilkerson and Olson (1997)، وبلوكير Plucker(1997) (١٩٩٧)، ومينون Menon(1993)، وهاركوم

Harcum(1989) تضم أطيافاً واسعة من المشكلات، ممّا يدعو إلى الحاجة للخروج برؤيا واضحة ومحدّدة عن مشكلات الدلالة الإحصائية.

١٠ - هناك تباين في توجهات الدراسات السابقة في استراتيجيات الحل، هل هي نقض الاستخدام الحالي للدلالة الإحصائية مثل دراسة مينون (Menon(1993، أم تطوير الموجود مثل ما ذهبت إليه دراسات ليفين (Levin(1998، وثومبسون Thompson(1999)، وماكلين (Mclain(1995، وكيرك Kirk(1996).

١١ - هناك دعوات في بعض الدراسات مثل ما دعت إليه دراسة فاقلي و ماك كنّي(Fagley and McKinny(1983 لأن تكون الدلالة الإحصائية جزء من مجموعة معايير، لكنّ هذه المعايير ومكانة الدلالة الإحصائية فيها غير واضحة.

١٢ - جميع الدراسات التطبيقية المتضمنة في الدراسات السابقة مثل: هوبارد وريان Hubbard and Ryan(2000)، وميتاق (Mittag(1999، وثومبسون Thompson(1999)، وسنايدر وثومبسون (Thompson(1998، وثومبسون وسنايدر (Snyder(1998 & Thompson، وثومبسون وسنايدر Thompson & Snyder(1997)، وويلكرسون وأولسون Wilkerson and Olson (1997)، وبلوكير (Plucker(1997 (١٩٩٧)، وكيرك Kirk(1996)، وماكلين (Mclain(1995، وهاركوم (Harcum(1989، وفاقلي وماك كنّي(Fagley and McKinny(1983 أفادت بوجود مشكلات في استخدام الدلالة الإحصائية في البحث التربوي على المستوى العالمي وهناك حاجة لمعرفة مدى وجود تلك المشكلات في البحث التربوي المحلّي.

١٣ - مشكلات الدلالة الإحصائية في البحث التربوي لم تحظ بنفس القدر من الاهتمام في الدراسات العربية مقارنة بالدراسات الأجنبية.

من خلال استعراض الملاحظات السابقة يرى الباحث بأنّ الدراسة الحالية يمكن أن تقدّم معالجة جديدة لمشكلات الدلالة الإحصائية على النحو التالي:

١ - الخروج بصورة شاملة ودقيقة عن أنواع وحجم مشكلات الدلالة الإحصائية في البحث التربوي، وليس بأشكال مجتزأة كما هي في الأبحاث السابقة.

٢ - التركيز على توضيح المفاهيم التي تساعد في دقّة الممارسة وإزالة اللبس وسوء التفسير.

٣ - بلورة أساليب محدّدة، ومحاولة صياغتها بشكل إجرائي تساعد الباحثين مستقبلاً في اختبار الفرضيات، وحل مشكلات الدلالة الإحصائية.

٤ - أنّ هذه الدراسة مخصّصة لاستيعاب مشكلات الدلالة الإحصائية، وتقديم حلول للمشكلات، ولا تتناول مشكلات الدلالة الإحصائية كمشكلات جانبية مع مشكلات أخرى في البحث التربوي.

٥ - تكشف هذه الدراسة عن مستوى وجود تلك المشكلات في البحث التربوي المحلي والمتمثّل برسائل الماجستير بكلية التربية بجامعة أم القرى بمكة المكرمة، وهي بذلك تتجاوز نظيراتها من الدراسات العربية السابقة حيث أنّها تركز على المشكلات الجوهرية التي تعتمد على سوء فهم معنى نتائج اختبارات الدلالة الإحصائية في البحث التربوي حتى مع افتراض أنّ الباحث قام بجميع الإجراءات السليمة في استخدام الأساليب الصحيحة والمناسبة والمتحققة بالافتراضات والاشتراطات اللازمة.

الفصل الثالث

إجراءات الدراسة

منهج الدراسة

استخدم الباحث للإجابة عن التساؤلات الثلاثة الأولى في هذه الدراسة منهج تحليل المعلومات، وهو أحد نماذج الدراسات المسحية التي تعد أحد الأنماط الأساسية في منهج البحث الوصفي، حيث يرتبط تحليل المعلومات (الوثائق) ارتباطاً وثيقاً بالبحوث التاريخية، فكلا المنهجين: منهج تحليل المعلومات الوصفي والمنهج التاريخي يفحصان السجلات، إلاّ أنّ البحوث التاريخية تهتم أساساً بالماضي البعيد بينما تتعلق البحوث الوصفية بالوضع الراهن(فان دالين، ١٩٦٩، ص:٣٤٠ -٣٤٨).

ولا يقتصر المنهج الوصفي على وصف الظاهرة وجمع البيانات فيها، بل لابد من تصنيف هذه المعلومات وتنظيمها وتلخيصها بعناية، ثمّ تحليل تلك المعلومات والأدلة بعمق في محاولة لاستخلاص تعميمات ذات معنى ومغزى تؤدي إلى تقدّم المعرفة (عبيدات، ١٩٩٩ :ص٢٤٨).

وحتى مع تفضيل كثير من النّاس بعض أنواع البحوث التطبيقية الأخرى التي تقدّم نتائج عملية مباشرة، وتفضيلها على البحوث الأساسية البحتة التي تجاهد لكشف أسرار الطبيعة العلمية الهامة، فإنّ هذا المنهج الوصفي لتحليل المعلومات والذي هو بحث أساسي بحت قد يؤدي في النهاية إلى نتائج أكثر فائدة من البحوث التطبيقية، المشكلة أنّ هذه البحوث البحتة لا يستطيع أن يدرك بعض النّاس كيف تستطيع أن تحسّن حياتهم من خلال تطوير العلم والمعرفة، مع أنّ هذه الدراسة الأساسية البحتة قد تؤدي إلى اكتشافات لا تقتصر على تحسين الممارسات والتطبيقات في مجال واحد محدّد مثلما تصل إليه الدراسة التطبيقية، بل قد يمتد أثر تلك الدراسة البحتة إلى مجال أوسع وأثر أكبر (فان دالين، ١٩٦٩ :ص١٣).

ويستطيع منهج تحليل المعلومات الوصفي المعتمد في هذه الدراسة أن يصف ظروفاً وممارسات معينة توجد في الواقع، وأن يبرز الاتجاهات، وأن يكشف نواحي الضعف، وأن يتتبع تطوّر أعمال ومفاهيم معينة، وأن يقيّم العلاقات بين الأهداف المرسومة لبعض الإجراءات والممارسات، وأن يميط اللثام عن التحيّزات، وغيرها من مميزات هذا المنهج (فان دالين، ١٩٦٩ : ص٣٤٩). وإذا لم يقم الباحثون بجمع حقائق متعلقة عن الموقف، فإنّهم لن يستطيعوا الإحساس بما له أهمية أو دلالة، وإذا لم يملكوا الحقائق، فإنّهم لن يستطيعوا ملاءمتها داخل الأطر المعقدة من العلاقات التي تكشف عن نظريات وقوانين علمية عامة، لتنتج مستوى أكثر تقدماً من المعرفة (فان دالين، ١٩٦٩ : ص٣٨٣).

ولا يخلو هذا المنهج كغيره من المناهج العلمية من قصور. فإنّ مشكلة هذا المنهج كما يراها فان دالين(١٩٦٩) تتمثّل في عيبين أساسين هما: الأول إمكانية اشتقاق الباحثين

استنتاجات خاطئة من المعلومات، والثاني اخفاقها في تحليل مدى الثقة بالمواد المصدرية (ص٣٥٠ -٣٥١).

المشكلة الأولى لهذه الطريقة لا يمكن التخلّص منها تماماً. فمع بذل الباحثين قصارى جهدهم العلمي والذهني في الوصول إلى استنتاجات علمية صحيحة، فإنّ احتمالية الوقوع في أخطاء غير مقصودة واردة على الجهد الإنساني. يجب أن يقوم الباحث بالجهد العلمي والأخلاقي الذي يفترض أن يكون خليقاً به. ويجب على الباحث أن يستنفذ سعته وطاقته، وأن يبذل قصارى جهده في التحرّي. وبعد ذلك تأتي إجراءات أخرى مثل عرض الدراسة على محكمين ومناقشين، ونشر الأعمال العلمية في أوساط المتخصصين، لتكون كفيلة بتمحيص ذلك الجهد العلمي. وكما ذكر ثومبسون Thompson(1995) بأنّ بعض التقييمات تخضع لقيم الإنسان ولا يمكن للرياضيات أو أي تقنيات أخرى أن توظف لهروب الإنسان من مسئوليته الموجودة لاتخاذ الأحكام، وسواء أعجبك هذا أم لا، العلم الامبريقي عمل ذاتي لا مفر منه.

أمّا العيب الثاني لهذا المنهج، فقد حاول الباحث التقليل من تأثيره من خلال كثرة البحث في المراجع العلمية ذات العلاقة بالمشكلة، والاعتماد على مراجع علمية قوية منشورة في مجلّات علمية عريقة، أو صادرة من مؤسسات مرموقة، كما أنّ الباحث اعتمد على آخر ما كتب في أدبيات المشكلة، ويدل على ذلك أنّ غالبية مراجع الدراسة حديثة جداً.

وللإجابة عن التساؤلين الرابع والخامس استخدم الباحث المنهج التقويمي Evaluation Research، وهو أيضاً يعتمد على تجميع المعلومات والحقائق ومقارنتها وتحليلها وتفسيرها للوصول إلى تفسيرات واستنتاجات واقعية قابلة للتطبيق والتطوير (بدر، ١٩٨٤: ص ٢٥).

أداة الدراسة

قام الباحث وعلى ضوء الدراسات المتصلة بموضوع البحث بإعداد أداة الدراسة وهي عبارة عن استمارة خاصة لجمع البيانات من عينة مجتمع الدراسة. وقد اعتمد الباحث لتصميم الأداة الخطوات التالية:

١ - قراءة أحدث ماكتب في أدبيات الموضوع.

٢ - تحديد أهم مشكلات الدلالة الإحصائية في البحث التربوي.

٣ - تحديد أهم الأساليب الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في البحث التربوي.

٤ - إعداد الأداة بصورتها المبدئية بناء على قراءات الباحث في موضوع الدراسة وذلك في مرحلة إعداد خطة الدراسة. انظر النموذج بالملحق رقم (١).

٥ - قام الباحث وبعد الإجابة على تساؤلات الدراسة الثلاثة الأولى وعرض الأداة على بعض المحكمين بإعادة بناء الأداة بشكلها النهائي حسب النموذج بالملحق رقم (٢).

مجتمع الدراسة

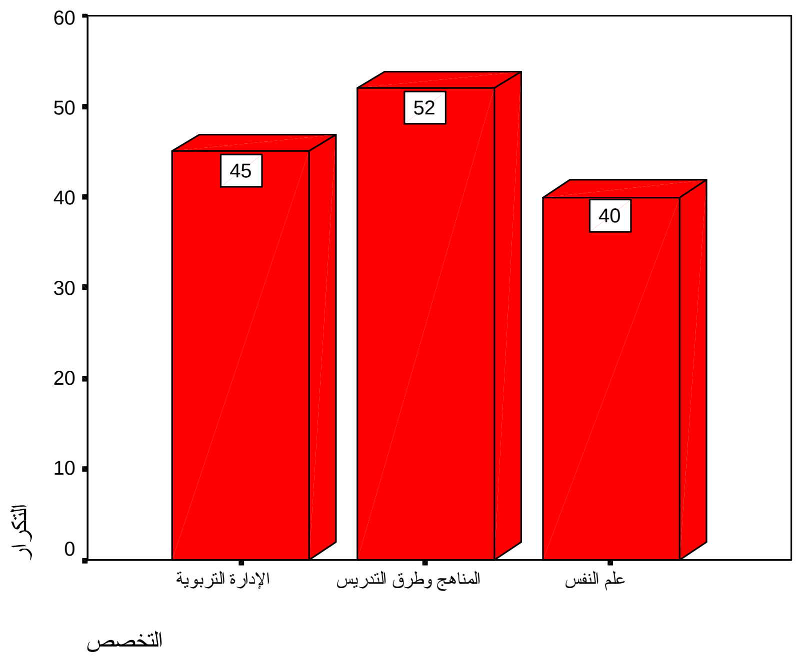
المجتمع الأصلي هو عدد رسائل الماجستير المقدمة لكلية التربية في الأقسام التربوية الثلاثة (الإدارة والتخطيط، المناهج وطرق التدريس، وعلم النفس) المقدمة خلال السنوات الثلاث (١٤١٧، ١٤١٨، ١٤١٩هـ)، وكان عددها حسب نظام التصنيف والإيداع بمكتبة الملك فهد الوطنية بالرياض (١٣٧) رسالة علمية.

وقد توزع المجتمع الإحصائي لهذه الدراسة على الأقسام على النحو التالي:

جدول ٣ : توزيع مجتمع الدراسة على أقسام كلية التربية الثلاثة

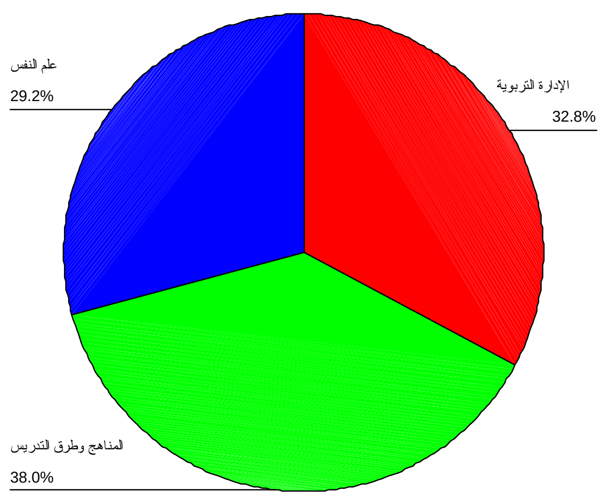
القسم	التكرارات	النسبة المئوية
الإدارة التربوية والتخطيط	٤٥	٪٣٢,٨
المناهج وطرق التدريس	٥٢	٪٣٨,٠
علم النفس	٤٠	٪٢٩,٢
المجموع	١٣٧	٪١٠٠

من الجدول السابق يتضح أنّ أكبر عدد لرسائل الماجستير العلمية الموجودة في المجتمع الإحصائي لهذه الدراسة هي الرسائل المقدمة لقسم المناهج وطرق التدريس (٥٢) رسالة، وأنّ أقلها هي الرسائل المقدمة لقسم علم النفس (٤٠) رسالة. الرسم البياني التالي يوضح توزع أعداد الرسائل في مجتمع الدراسة حسب أقسام كلية التربية:



رسم بياني ١ : توزيع أعداد الرسائل في مجتمع الدراسة حسب أقسام كلية التربية

يمثّل الرسم البياني التالي نسب رسائل الماجستير في مجتمع الدراسة حسب أقسام كلية التربية:



رسم بياني ٢ : نسب رسائل الماجستير في مجتمع الدراسة حسب أقسام كلية التربية

يتضح من الرسم البياني السابق أنّ نسب رسائل الماجستير في المجتمع الإحصائي حسب الأقسام هي على النحو التالي:

(١) قسم المناهج وطرق التدريس (٣٨,٠٪)

(٢) قسم الإدارة التربوية والتخطيط (٣٢,٨٪)

(٣) قسم علم النفس (٢٩,٢٪).

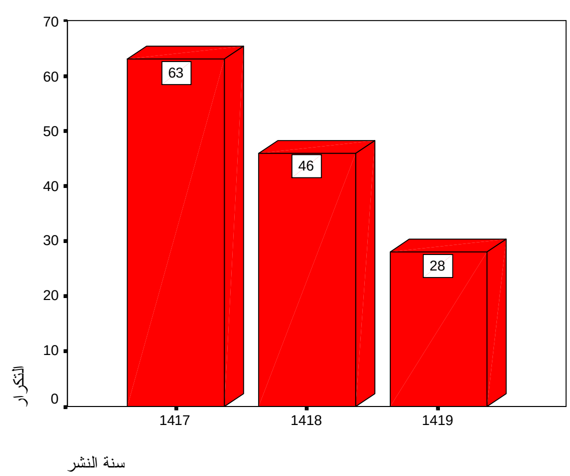
أمّا توزيع الرسائل في مجتمع الدراسة حسب سنوات النشر فيوضحه الجدول (٤):

جدول ٤ : توزيع مجتمع الدراسة حسب سنوات النشر

السـنة	التكـرارات	النسبة المئوية
١٤١٧ هـ	٦٣	٤٦,٠٪
١٤١٨ هـ	٤٦	٣٣,٦٪
١٤١٩ هـ	٢٨	٢٠,٤٪
المجموع	١٣٧	١٠٠٪

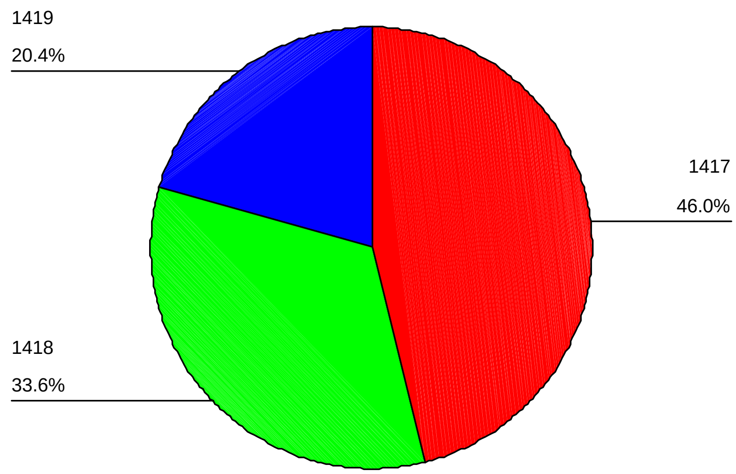
من الجدول السابق يتضح أنّ أكبر عدد من رسائل الماجستير في مجتمع الدراسة هي المنشورة عام ١٤١٧هـ (٦٣) رسالة، وأقلها هي المنشورة عام ١٤١٩هـ (٢٨) رسالة.

الرسم البياني التالي يوضح توزيع أعداد رسائل الماجستير في مجتمع الدراسة حسب سنوات النشر الثلاث.



رسم بياني ٣ : توزيع أعداد رسائل الماجستير في مجتمع الدراسة حسب سنوات النشر

ويوضح الرسم البياني التالي توزيع نسب رسائل الماجستير في مجتمع الدراسة حسب سنوات النشر:



رسم بياني ٤ : نسب توزيع رسائل الماجستير في مجتمع الدراسة حسب سنوات النشر

يتضح من الرسم البياني السابق، أنّ نسب رسائل الماجستير في المجتمع الإحصائي حسب سنوات النشر هي على النحو التالي: (١) الرسائل المنشورة عام ١٤١٧هـ (٤٦,٠٪)، (٢) الرسائل المنشورة عام ١٤١٨هـ (٣٣,٦٪) (٣) الرسائل المنشورة عام ١٤١٩هـ (٢٠,٤٪).

ولتحديد عينة البحث القصدية من هذا المجتمع الإحصائي، قام الباحث بالاطلاع على جميع رسائل الماجستير في الأقسام الثلاثة بكلية التربية لتحديد علاقتها بعينة الدراسة حيث كانت الرسائل في المجتمع الإحصائي على النحو التالي:

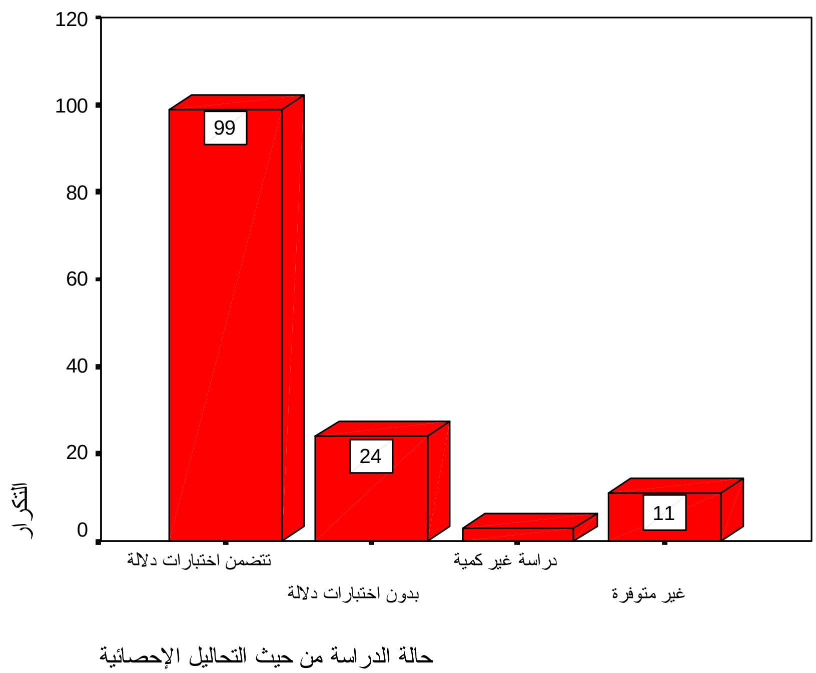
جدول ٥ : توصيف المجتمع الإحصائي حسب أنواع التحاليل الإحصائية المستخدمة

حـالـة الدراسـة	التكرارات	النسبة المئوية
استخدمت اختبارات دلالة إحصائية	٩٩	٪٧٢,٣
لم تستخدم اختبارات الدلالة الإحصائية	٢٤	٪١٧,٥
دراسة غير كمية (لم تستخدم أساليب إحصائية)	٣	٪٢,٢
غير متوفرة	١١	٪٨,٠
المجموع	١٣٧	٪١٠٠

من الجدول السابق يتضح أنّ عدد الرسائل التي تتضمن اختبارات دلالة إحصائية والمتوفرة بمكتبة الملك فهد الوطنية هي (٩٩) رسالة ماجستير، وهي التي ستمثّل العينة القصدية لهذه الدراسة، بينما يوجد عدد (٢٤) رسالة ماجستير استعملت أساليب إحصائية غير

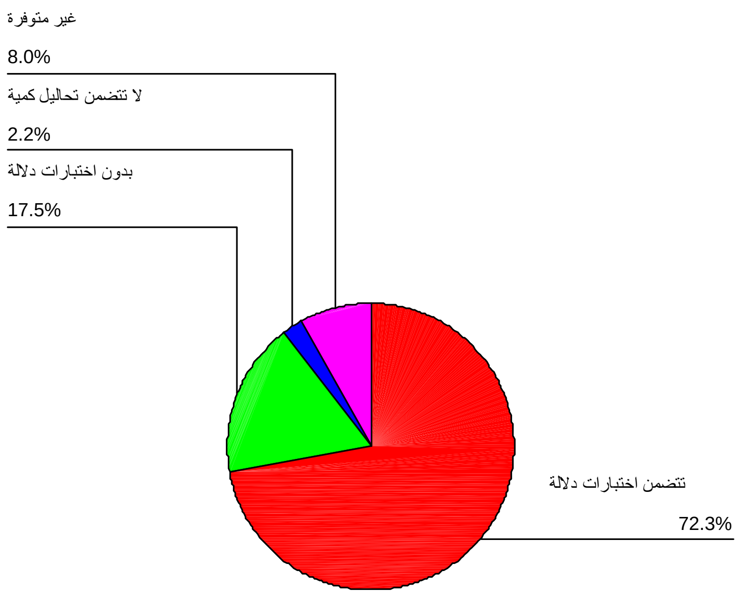
اختبارات الدلالة الإحصائية، أمّا عدد (٣) رسالة ماجستير من المجتمع الإحصائي لهذه الدراسة فلم تستخدم الأساليب الإحصائية الكمية على الاطلاق، بينما كان عدد الرسائل المسجلة ضمن سجلات مكتبة الملك فهد الوطنية والتي لم تتوفّر أثناء دراسة الباحث هي (١١) رسالة.

الرسم البياني التالي يوضح تمثيل أعداد رسائل الماجستير العلمية بالمجتمع الإحصائي للدراسة حسب حالتها وعلاقتها بعينة الدراسة:



رسم بياني ٥ : أعداد رسائل الماجستير في المجتمع الإحصائي حسب أنواع التحاليل الإحصائية المستخدمة

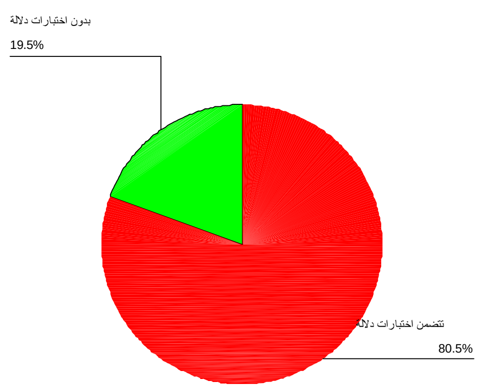
الرسم البياني التالي يوضح النسب المئوية لكل حالة من حالات الرسائل في المجتمع الإحصائي حسب علاقتها بعينة الدراسة:



رسم بياني ٦ : النسب المئوية لرسائل الماجستير في المجتمع الإحصائي حسب أنواع التحاليل الإحصائية المستخدمة

يتضح من الرسم البياني السابق أنّ أعلى نسبة من رسائل المجتمع الإحصائي هي الرسائل التي استخدمت اختبارات الدلالة الإحصائية (٧٢,٣٪) ، بينما (١٧,٥٪) من رسائل المجتمع الإحصائي استخدمت أساليب إحصائية غير اختبارات الدلالة الإحصائية، أمّا (٢,٢٪) من الرسائل في المجتمع الإحصائي فلم تستعمل أصلاً أي أسلوب إحصائي كمي، أمّا (٨,٠٪) من رسائل المجتمع الإحصائي فلم تتوفر في مكتبة الملك فهد الوطنية وقت إجراء الدراسة.

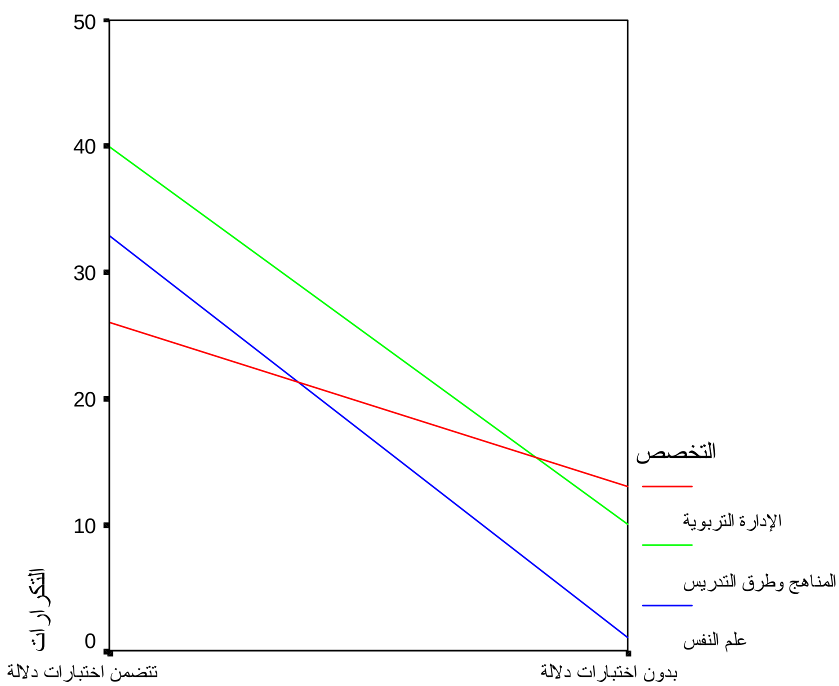
كما يوضح الرسم البياني التالي نسبة اعتماد رسائل الماجستير في مجتمع الدراسة على أساليب الدلالة الإحصائية مقارنة بغيرها من الأساليب :



رسم بياني ٧ : النسب المئوية للرسائل التي استخدمت اختبارات الدلالة الإحصائية وتلك التي لم تستخدمها

من الرسم البياني السابق يتضح أنّ: عدد رسائل الماجستير بكلية التربية بجامعة أم القرى التي استخدمت اختبارات الدلالة الإحصائية بلغت نسـبتها (٨٠,٥٪) من مجموع رسائل الماجستير بكلية التربية التي استخدمت أساليب إحصائية. ويتضح من المشاهدة السابقة أنّ رسائل الماجستير بكلية التربية بجامعة أم القرى تعتمد بشكل كبير على اختبارات الدلالة الإحصائية في تحليل النتائج، وهذا يؤكد ما توصلت إليه الدراسات السابقة من أنّ البحث في العلوم الإنسانية يعتمد حالياً على اختبارات الدلالة الإحصائية، وهذا أيضاً يمنح هذه الدراسة أهمية أكبر لأنّها تشخص مشكلات الأساليب الإحصائية الأكثر استخداماً في رسائل الماجستير بكلية التربية بجامعة أم القرى.

كما يبيّن الرسم البياني التالي الفروق بين أقسام كلية التربية في مجتمع الدراسة في درجة اعتمادها على اختبارات الدلالة الإحصائية.



رسم بياني ٨ : مقارنة الفروق في بين أقسام كلية التربية في تكرارات استخدام وعدم استخدام اختبارات الدلالة الإحصائية

يتضح من الرسم البياني السابق أنّ أعلى فرق في نسبة الاعتماد على اختبارات الدلالة الإحصائية مقارنة بالأساليب الإحصائية الأخرى هي في قسم علم النفس، يليه قسم المناهج وطرق التدريس، ثمّ قسم الإدارة التربوية والتخطيط.

عينة الدراسة

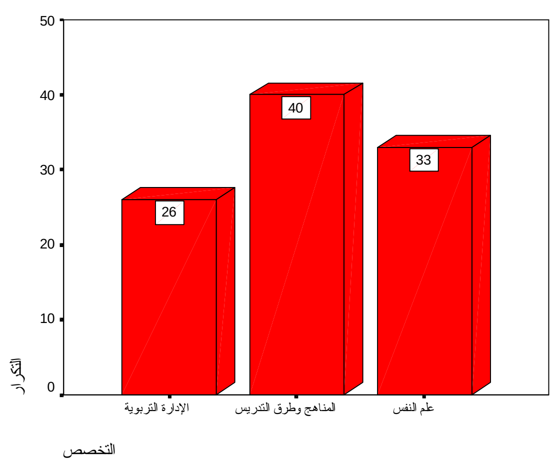
من نتيجة اطلاع الباحث على جميع رسائل المجتمع الإحصائي المتوفرة بمكتبة الملك فهد الوطنية، والتي توضحها الرسوم البيانية السابقة يتضح أنّ عدد عينة هذه الدراسة التي تمثّل جميع رسائل الماجستير المقدمة للأقسام الثلاثة (الإدارة التربوية والتخطيط، المناهج وطرق التدريس، علم النفس) في كلية التربية بجامعة أم القرى خلال الأعوام (١٤١٧ ، ١٤١٨، ١٤١٩هـ) والتي استخدمت اختبارات الدلالة الإحصائية هي (٩٩) رسالة علمية.

وقد اختار الباحث هذه العينة القصدية لأنّها تمثّل أحدث ما تمت إجازتها من رسائل الماجستير بكلية التربية بجامعة أم القرى، وهي بذلك تمثّل تطوّر برامج الماجستير ونضجها في كلية التربية بجامعة أم القرى مع مرور الأعوام، ولم تتجاوز أحدث سنة للنشر عام ١٤١٩هـ بسبب أنّ تلك الرسائل لم تكتمل بعد في مكتبة الملك فهد الوطنية بالرياض وذلك حسب إفادة الموظفين المختصين.

الجدول والرسم البياني التاليين يوضحان أعداد عينة الدراسة حسب الأقسام الثلاثة:

جدول ٦ : أعداد عينة الدراسة حسب أقسام كلية التربية

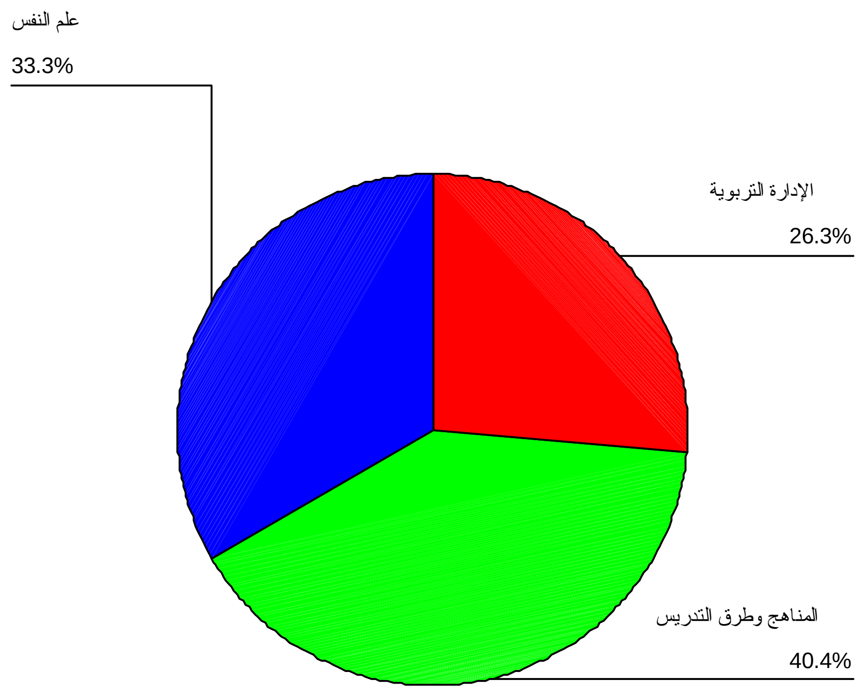
القسم	التكرارات	النسبة المئوية
الإدارة التربوية والتخطيط	٢٦	٪٢٦,٣
المناهج وطرق التدريس	٤٠	٪٤٠,٤
علم النفس	٣٣	٪٣٣,٣
المجموع	٩٩	٪١٠٠



رسم بياني ٩ : أعداد عينة الدراسة حسب أقسام كلية التربية

يتضح من الجدول والرسم البياني السابقين أنّ أكبر عدد لرسائل الماجستير في عينة الدراسة هي المقدمة لقسم المناهج وطرق التدريس (٤٠ رسالة)، تليها الرسائل المقدمة لقسم علم النفس (٣٣ رسالة)، وأخيراً الرسائل المقدمة لقسم الإدارة التربوية والتخطيط (٢٦ رسالة).

الرسم البياني التالي يوضح النسبة المئوية لرسائل كل قسم من الأقسام الثلاثة في عينة الدراسة:



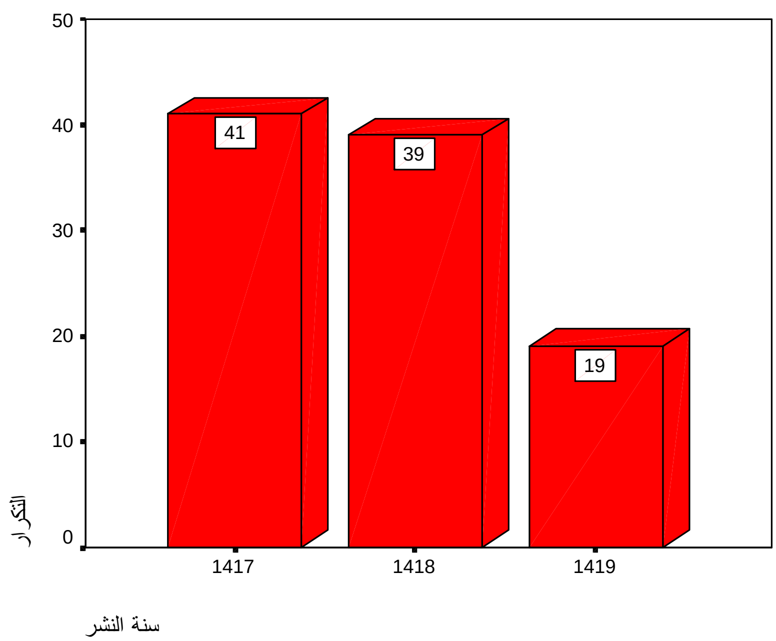
رسم بياني ١٠ : النسبة المئوية لرسائل كل قسم من أقسام كلية التربية في عينة الدراسة

من الرسم البياني السابق يتضح أنّ النسبة المئوية لرسائل كل قسم من الأقسام الثلاثة في كل التربية هي: قسم المناهج وطرق التدريس (٤٠,٤٪)، يليه قسم علم النفس (٣٣,٣٪)، وأخيراً قسم الادارة التربوية والتخطيط (٢٦,٣٪).

الجدول والرسم البياني التاليان يوضحان أعداد الرسائل في عينة الدراسة حسب سنوات النشر:

جدول ٧ : أعداد رسائل الماجستير في عينة الدراسة حسب سنوات النشر

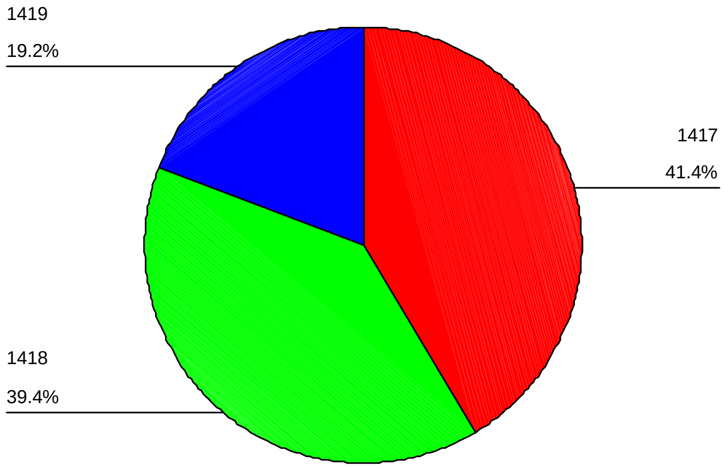
سـنـة النشــر	التكـرارات	النسبة المئوية
١٤١٧ هـ	٤١	٪٤١,٤
١٤١٨ هـ	٣٩	٪٣٩,٤
١٤١٩ هـ	١٩	٪١٩,٢
المجموع	٩٩	٪١٠٠



رسم بياني ١١ : أعداد رسائل الماجستير في عينة الدراسة حسب سنوات النشر

من الجدول والرسم البياني السابقين يتضح أنّ أكبر عدد للرسائل في عينة الدراسة هي الرسائل المنشورة عام ١٤١٧هـ (٤١ رسالة)، تليها الرسائل المنشورة عام ١٤١٨هـ (٣٩ رسالة)، وأخيراً الرسائل المنشورة عام ١٤١٩هـ (١٩ رسالة).

الرسم البياني التالي يوضح نسبة الرسائل المكونة لعينة الدراسة حسب سنوات نشرها:



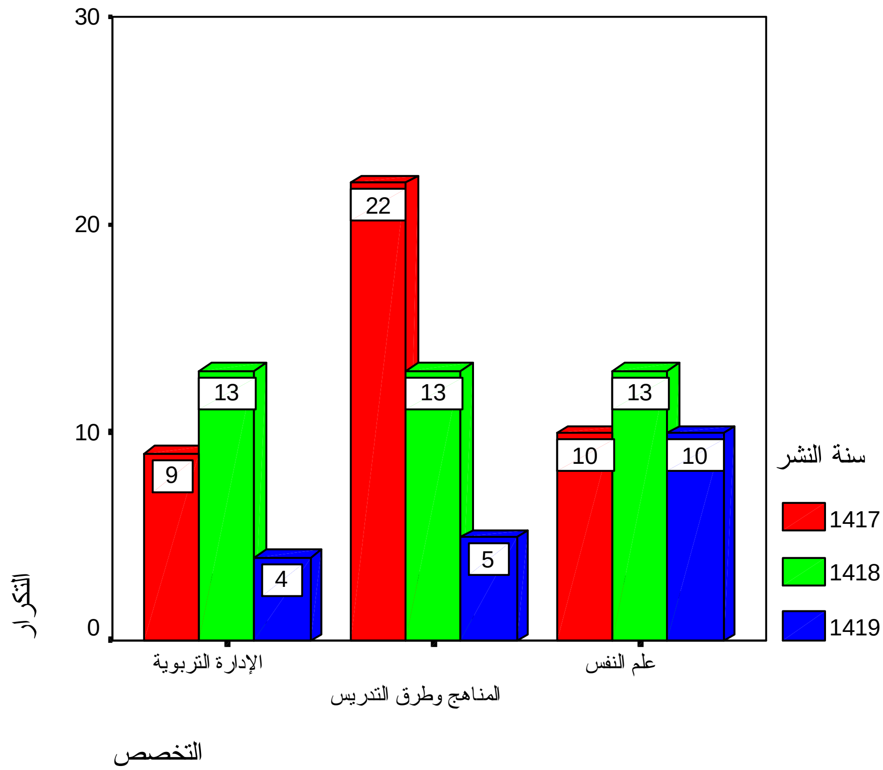
رسم بياني ١٢ : نسب الرسائل في عينة الدراسة حسب سنوات النشر

من الرسم البياني السابق يتضح أنّ أعلى نسبة للرسائل المكونة لعينة الدراسة هي المنشورة عام ١٤١٧هـ (٤١,٤٪)، وأقلها المنشــورة عام ١٤١٩هـ (١٩,٢٪).

الجدول والرسم البياني التاليان يوضحان توصيفاً لأعداد الرسائل في عينة الدراسة حسب التخصص وسنوات النشر:

جدول ٨ : أعداد الرسائل في عينة الدراسة حسب التخصصات وسنوات النشر

القسم	سنوات النشر			المجموع
	١٤١٧	١٤١٨	١٤١٩	
الإدارة التربوية والتخطيط	٩	١٣	٤	٢٦
المناهج وطرق التدريس	٢٢	١٣	٥	٤٠
علم النفس	١٠	١٣	١٠	٣٣
المجموع	٤١	٣٩	١٩	٩٩



رسم بياني ١٣: أعداد الرسائل في عينة الدراسة حسب التخصصات وسنوات النشر

يتضح من الجدول والرسم البياني السابقين أنّ أكبر عدد رسائل في تخصص الإدارة التربوية والتخطيط الممثلة في عينة الدراسة نشرت عام ١٤١٨هـ وأقلها نشرت عام ١٤١٩هـ، بينما أكبر عدد رسائل في تخصص المناهج وطرق التدريس والممثلة في عينة هذه الدراسة نشرت عام ١٤١٧هـ وأقلها نشرت عام ١٤١٩هـ. أمّا توزيع عدد الرسائل التي تخص قسم علم النفس والممثلة في عينة الدراسة فهي أكثر اعتدالية في التوزيع، حيث أنّ أكبر عدد من الرسائل نشرت عام ١٤١٨هـ، بينما تتوزع أعداد الرسائل المتبقية بالتساوي بين عامي ١٤١٧هـ و ١٤١٩هـ.

خطوات إجراء الدراسة

لتحقيق أهداف الدراسة، والاجابة عن تساؤلاتها، قام الباحث باتباع الخطوات التالية:

أ) **خطوات الإجابة على التساؤلات الثلاثة الأولى**:

١ - قام الباحث بالبحث المكثّف عمّا كتب عن الدلالة الإحصائية في العلوم الإنسانية، وقد استعمل الباحث عدّة قواعد بيانات منها: قاعدة بيانات إريك ERIC ، قاعدة بيانات مؤسسة الملك فيصل الخيرية، قاعدة بيانات مكتبة الملك فهد الوطنية، قواعد البيانات المتاحة لدى مدينة الملك عبد العزيز للعلوم والتقنية، إضافة لعدّة مواقع متخصصة على شبكة الإنترنت.

٢ - قام الباحث بجمع ما لا يقل عن (١٠٠) مادّة علمية كاملة Full Text ما بين مقال علمي منشور في مجلات، وأوراق علمية مقدمة في مؤتمرات، وتعليقات على مقالات، ورسائل علمية غير منشورة، إضافة إلى مواد علمية منشورة في مواقع متخصصة ومرموقة على شبكة الإنترنت.

٣ - قام الباحث بالإطلاع على جميع تلك المواد وتلخيص وتفريغ ما له علاقة بموضوع هذه الدراسة خاصة المشكلات والحلول البديلة وذلك في بطاقات الفهرسة المخصصة لأغراض البحث.

٤ - قام الباحث بعد ذلك بإعادة تصنيف تلك البطاقات حيث استخلص الباحث أهم مشكلات الدلالة الإحصائية وذلك بناء على معيار اختاره الباحث وهو الإشارة إلى المشكلة في ثلاث وثائق على الأقل من المواد العلمية التي جمعها الباحث.

٥ - قام الباحث ومن خلال ما تجمع لديه من مشكلات واقتراحات بصياغة استراتيجية مقترحة لتحسين تفسير نتائج اختبارات الدلالة الإحصائية في البحث التربوي.

٦ - أخيراً قام الباحث بعرض نتائج التساؤل الأول والثاني والثالث حول المشكلات والحلول البديلة والاستراتيجيات الجديدة في الفصل الرابع (تحليل النتائج وتفسيرها).

(ب) خطوات الإجابة على التساؤلين الرابع والخامس (الدراسة التطبيقية)

١ - حصر الرسائل الجامعية (درجة الماجستير فقط) المعتمدة من كلية التربية بجامعة أم القرى بمكة المكرمة، في تخصصات: الإدارة التربوية، والمناهج وطرق التدريس، وعلم النفس، والتي قدمت للأقسام خلال الأعوام: ١٤١٧، ١٤١٨، ١٤١٩هـ. وقد اعتمد الباحث في حصر هذه الرسائل على النظام الآلي المتوفر لدى مكتبة الملك فهد الوطنية بالرياض، حيث حصل الباحث على مخرج مطبوع بقائمة جميع الرسائل التي تحقق الشرط السابق، حيث تضمنت القائمة البيانات التالية:

- رقم الطلب: وهو الرقم التصنيفي الذي تعتمده مكتبة الملك فهد الوطنية في إيداع الرسائل الجامعية.
- العنوان: وهو العنوان الكامل للرسالة الجامعية.
- المسئولية: ويتضمن اسم معدّ الرسالة، واسم الأستاذ المشرف عليها.
- تاريخ نشره: وهو العام الهجري، وما يوافقه من العام الميلادي اللذان تم فيهما نشر الرسالة الجامعية.
- الموضوعات: وتتضمن كلمات أساسية كمفاتيح للتصنيف والبحث بالموضوع.
- وضع المادة: هل هي مفهرسة ومودعة بالمكتبة، أم أنّها في طريقها للإيداع.
- نوع المادة: توضح هل هي رسالة ما جستير أم دكتوراة.
- موقع النسخة: لتحديد موقع وجودها في مكتبة الملك فهد الوطنية.

٢ - قام الباحث بمنح رقم تسلسلي لجميع رسائل المجتمع الإحصائي من الرقم (١) وحتى الرقم (١٣٧)، وذلك لتسهيل إجراءات ادخال أرقام الحالات في الحاسب الآلي، وكذلك لتسهيل مراجعة استمارة كل حالة.

٣ - قام الباحث بالاطلاع على جميع الرسائل بالمجتمع الإحصائي لهذه الدراسة، وتعبئة نموذج أداة الدراسة المعدّة لهذا الغرض. وقد كان إجراء الاطلاع على الرسائل يعتمد أولاً على مراجعة القسم الخاص باجراءات الدراسة الذي يتضمن الاختبارات الإحصائية المستخدمة. في حالة عدم تضمين أي

اختبارات دلالة إحصائية أو الإشارة إلى ذلك، قام الباحث بمراجعة سريعة لكامل فصل (تحليل النتائج) وذلك للتأكد من خلو الرسالة فعلاً من أي اختبارات دلالة إحصائية.

٤ - إذا كانت الرسالة لا تستخدم الأساليب الإحصائية، أو تستخدم أساليب إحصائية خلاف اختبارات الدلالة الإحصائية، يقوم الباحث بالتأشير على الخانة المعدّة لهذا الغرض في أداة البحث، بعد استكمال الخانات السابقة لها، ويوقف الباحث تعبئة بقية خانات الأداة.

٥ - إذا كانت الرسالة تستخدم اختبارات الدلالة الإحصائية، يقوم الباحث باستكمال تعبئة أداة الدراسة حسب النموذج في الملحق رقم (٢).

٦ - قام الباحث بعرض نتائج التساؤل الرابع والخامس المتعلقين بواقع مشكلات الدلالة الإحصائية والأساليب الإحصائية التي يمكن أن تقدّم حلولاً لتلك المشكلات في رسائل الماجستير بكلية التربية بجامعة أم القرى في فصل تحليل النتائج وتفسيرها.

الأساليب الإحصائية

- الأساليب الإحصائية الوصفية المتمثلة بحساب عدد التكرارات ، والمتوسطات، والنسب المئوية.
- الرسوم البيانية الموضحة لعدد التكرارات.
- الرسوم البيانية التي توضح النسب المئوية لتكرارات بعض الحالات بالنسبة للعدد الكلي لحالات العينة.
- الرسوم البيانية التي تمثّل أعمدة الخطأ وذلك لتمثيل حدود الثقة.
- حساب تقدير حدود الثقة حسب المعادلة العامة التالية:

حدود الثقة = الإحصاءة ± القيمة الحرجة × الخطأ المعياري

حيث يحسب الخطأ المعياري $s_{\bar{x}}$ كالآتي: $s_{\bar{x}} = \frac{s}{\sqrt{n}}$ ، وحيث تحدّد القيمة الحرجة حسب جدول التوزيع الطبيعي في كتب الإحصاء الذي يستخدم للحصول على القيم الحرجة المتعلقة بمستوى الثقة المرغوبة (Wilkinson,1992).

الفصل الرابع

تحليل النتائج ومناقشتها

تمهيد

تتضمن هذه الدراسة خمسة تساؤلات رئيسة، حاول الباحث من خلال هذا الفصل الإجابة عنها وهي:

١. ما أهم مشكلات الدلالة الإحصائية في البحث التربوي ؟

٢. ما أهم المفاهيم الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في البحث التربوي؟

٣. ما الأساليب الجديدة لحل مشكلات الدلالة الإحصائية في البحث التربوي؟

٤. ما واقع مشكلات الدلالة الإحصائية في رسائل درجة الماجستير بكلية التربية بجامعة أم القرى بمكة المكرمة؟

٥. ما مدى وجود المفاهيم الإحصائية التي يمكن أن تقدّم حلولاً لمشكلات الدلالة الإحصائية في رسائل درجة الماجستير بكلية التربية بجامعة أم القرى بمكة المكرمة؟

وقد قام الباحث بعرض ما توصل إليه من إجابات لهذه التساؤلات في هذا الفصل على جزأين: الأول للإجابة عن التساؤلات الثلاثة الأولى، والثاني للإجابة عن التساؤلين الأخيرين.

أولاً

إجابة الأسئلة

الأول والثاني والثالث

إجابة السؤال الأول

أهم مشكلات الدلالة الإحصائية

مشكلات الدلالة الإحصائية التي توصل إليها الباحث في هذا الفصل يمكن أن تقع جميعها مع افتراض أنّ البيانات أتت من تصميم يتضمن العشوائية، سواء عشوائية المعاينة أو التعيين. يجب على الباحث الذي يستخدم تصاميم مقارنة، واستراتيجيات التحليل الكمي للوصول إلى نتائج البحث أن يتأكد من أنّ النتائج التي يعرضها مفسّرة بطريقة صحيحة، فالتفسير غير الصحيح يقود إلى خسارة كبيرة في عدّة قطاعات، بما فيها التربية، وعلم النفس، والإدارة، والصحة،... (Kellow,1998). ومن خلال ما استعرضه الباحث في الإطار النظري لهذه الدراسة لمفهوم الدلالة الإحصائية، والعوامل المؤثرة فيها، ومن خلال دراسة الباحث لكثير ممّا كتب عن الدلالة الإحصائية ومشكلاتها، توصل الباحث إلى تحديد أهم مشكلات الدلالة الإحصائية في البحث التربوي على النحو التالي:

أولاً- استخدام نتائج الدلالة الإحصائية كتفسير لأهمية النتائج

يستخدم الكثير من الباحثين وبطريقة غير صحيحة الدلالة الإحصائية لتوضيح أهمية نتائجهم (Daniel,1997). وقد عبّر عن هذه المشكلة العديد من الباحثين مثل: دانييل Daniel(1997) وثومبسون Thompson(1988) وويلكنسون Wilkinson(1992). فقد قال دانييل Daniel(1997) : "الاعتراض الرئيس في تفسير اختبارات الدلالة الإحصائية، هو الافتراض بأنّ النتيجة الدالة إحصائياً من الضروري أن تكون نتيجة رائعة وجديرة بالملاحظة والاهتمام "(p.2). وقد اعتبر ثومبسون Thompson(1988) هذه الممارسة واحدة من أكثر الأخطاء الشائعة المعمول بها من قبل الباحثين في الدراسات العليا. وأكد هايز بأنّ الدلالة الإحصائية ليست أكثر من جملة تتعلق باحتمالية بعض النتائج المشاهدة و أنّها لاتضمن أهمية أو قيمة النتائج (Wilkinson,1992). ولعلّ أفضل من وضّح الخطأ في تفسير انخفاض قيمة الاحتمالية (r) كمؤشر على أهمية نتائج البحث، هو شيفر Shaver من خلال الحوار التمثيلي التالي (Thompson,1994):

كريس Chris: لقد وضعت مستوى الدلالة عند ٠,٠٥ كما اقترح عليّ مشرف الرسالة. ولذا فإنّ الفرق الكبير الذي ربما يحدث بالصدفة سيكون أقل من ٥٪ إذا كانت المجموعتين في الحقيقة ليست مختلفتين، هذه النتيجة غير المحتملة بالتأكيد يجب أن تكون مهمة.

جين Jean: انتظر لحظة، وتذكر اليوم الذي ذهبت فيه إلى المكتب للاتصال بالمنزل وبمجرد إنهائك لضغط الأرقام رفع ابنك الصغير السماعة للاتصال بشخص آخر وحدث الاتصال بينك وبينه وتحدث كل واحد منكم للآخر دون أن يرنّ الهاتف، إذاً هل في الحقيقة يجب أن يكون هذا الحدث مهماً! النتيجة غير المحتملة ليس بالضرورة أن تكون نتيجة مهمة.

وقد أصبح الباحثون في الغالب ينظرون لاختبار الدلالة الإحصائية كهدف رئيس في البحث، كما عبّر عن ذلك ياتس بقوله: "الباحثون أقصى ما يسعون له في البحث هو هل النتائج دالة أو غير دالة ؟ وتلك نهاية الأمر" (في: Wilkinson,1992,p.820) . ويؤكد ذلك شيفر Shaver(1992) حيث يرى أنّه من سوء الحظ، أن يعمل مؤلفو كتب الإحصاء وتقارير الأبحاث وبشكل متكرر تقريرات عن رفض الفرضية الصفرية H_o اعتماداً على النتيجة الدالة إحصائياً كأنّ ذلك هو مطلق استنتاجهم، وأنّ المتمعّن في مجلات البحث التربوي وكتب الإحصاء التربوي والنفسي واطروحات الدكتوراة سوف يتأكد بأنّ اختبارات الدلالة الإحصائية مستمرة في الهيمنة على تفسير البيانات الكمية في البحث التربوي. وكما قال ثومبسون Thompson(1995a) بأنّه عندما تتحقق النتيجة الدالة فإنّ المحلّل عادة ما يشعر بالثقة لإنهاء تحليله.

ويرى كيلو Kellow(1998) أنّ تفسير النتائج عندما تعتمد فقط على اختبارات الدلالة الإحصائية فإنّها تحجب وجود بعض التأثيرات القيّمة. كما يرى شيفر Shaver(1992) أنّ سيطرة وهيمنة اختبارات الدلالة الإحصائية هو اختلال وظيفي، لأنّ مثل هذه الاختبارات لا تزوّد بالمعلومات التي يفترض كثير من الباحثين أنّها تقدّمها، كما أنّ اختبار الدلالة الإحصائية يحرف الانتباه والجهد عن الاستراتيجيات الأكثر مناسبة مثل الإعادة والوعي بنتائج الدلالة العملية. وقد نقل ويلكنسون Wilkinson(1992) عن هايز رأيه بأنّ اختبار الدلالة الإحصائية هي أداة ولكنّها ليست حكماً على النوعية. كما يرى بعض الباحثين أنّ التركيز يفترض أن يكون على البيانات، وما إذا كانت البيانات تدعم الفرضية العلمية، وهي فكرة قال بها كارل بيرسون Karl Pearson عام ١٩٠١، وكذلك السير رونالد فيشر Ronald Fisher عام ١٩٢٥ حيث افترض فيشر أنّ الباحثين يلحقون اختبار الدلالة الإحصائية في تحليل التباين بنسبة الارتباط (Eta2) التي تقيس قوة التلازم بين المتغيّرات المستقلة والتابعة (Kirk,1996). وقد اعتبر هوستون (Huston,1995) أنّ العادة الراسخة في تعريف النتائج البحثية فقط من خلال مستويات الدلالة الإحصائية، تشوه تفسير النتائج، وذلك باستبعاد قيم معتبرة مثل تحليل القوة وقياسات مقدار التأثير. وقد ذكر كيلو Kellow(1988) بأنّه أصبح واضحاً أنّ تفسير النتائج بالاعتماد فقط على اختبارات الدلالة الإحصائية ربما يحجب التأثيرات الموجودة والكامنة. وقد لخص ماكلين وإرنست McLean

and Ernest(1998) هذه المشكلة بقولهما "الاعتراض الأهم الذي يركز عليه في أزمة الممارسات هو أنّ الباحثين يجب أن يقوّموا الأهمية العملية للنتائج وليس فقط الدلالة الإحصائية" (p.4).

وقد رأى بالوميرس Palomares(1990) بأنّ التركيز على الإحصاء وإجراءات اختبار الدلالة نتج عنه توجه منهج البحث نحو تأسيس تعميمات ذات تأثيرات مضرّة على التراكم العلمي للمعرفة في التربية. وقد وصف ليكن هذا الوضع بقوله: "قيمة أي بحث يمكن أن تحدّد ليس فقط من النتائج الإحصائية، ولكن من التقويم الماهر والموضوعي، ودرجة توظيف التحكّم التجريبي، وتقنيات القياس المتطورة، والأهمية العملية للظاهرة المدروسة، ... وهكذا... وكل ما سبق يشير إلى أنّ هناك أشياء كثيرة أكثر من مجرد اختبار للدلالة الإحصائية " (Keaster,1988,p.13).

لكنّ المشكلة لم تتوقف عند هذا الحد. بل ذهب بعض الباحثين إلى أبعد من ذلك، حيث أخذوا يفاضلون بين أهمية الدراسات حسب قيم الاحتماليات المحسوبة ($r_{calculated}$). وقد وجد كل من كاريج وأيزون وميتز في دراسة لهم عام ١٩٧٦ على عدّة دراسات في مجلات علم نفس مختارة، أنّ النتائج التي كانت فيها ($r < 0.001$) أعطيت قيمة أكبر من الدراسة التي فيها ($r < 0.05$)، وهذا يعني أنّ قيمة الدراسة البحثية مرتبطة بالمستويات المنخفضة من قيمة الدلالة (Wilkinson,1992).

هذه المشكلة التي بدأت تتضح في الميدان التربوي بشكل أكبر، دفعت لان Lane(2000) إلى التصريح في ثنايا كتابه عن إدراكه المتأخر لهذه المشكلة خاصة عندما يقلّل من أهمية بعض النتائج بسبب عدم وصولها إلى مستويات الدلالة الإحصائية، حيث قال بأنّه عندما كتب كتابه الأخير كان جاهزاً للتسليم بأنّ قيم (r) لها استخدام وفائدة عند وصف تأثيرات كثيرة، مثل عند عرض نتائج (٢٠) ارتباط في جدول، حيث يمكن وضع النجمات للإشارة إلى مستوى الدلالة الإحصائية، لكنّه الآن غير متأكد من فائدة هذه النجمات، لأنّ النتائج غير الدالة إحصائياً ربما تكون مهمة.

لقد اتضح من مناقشة العوامل المؤثرة في نتائج اختبارات الدلالة الإحصائية، أنّ هناك سبعة عوامل، أحدها – وللأسف ليس أكثرها تأثيراً – التأثيرات الفعلية في الدراسة، بينما هناك عوامل أخرى ليست ذات علاقة بالهدف الرئيس من اختبار الفرضيات مثل أحجام العينات شديدة التأثير في نتائج اختبارات الدلالة الإحصائية، وهذا يعني أنّه يمكن الحصول على نتائج ذات دلالة إحصائية، بغض النظر عن التأثير الفعلي للمعالجات في الدراسة، عن طريق التحكم في بعض العوامل السابقة.

ثانياً - استخدام نتائج الدلالة الإحصائية كتفسير لتأثير المعالجة

يستخدم الباحثون بعض اختبارات الدلالة الإحصائية واختبار الفرضيات لمحاولة التعرّف هل النتائج المشاهدة سببها المعالجة التجريبية، ومن ثمّ الوصول إلى القرار بتأثير المعالجة أو عدم تأثيرها. لكنّ للأسف نتيجة اختبار الدلالة الإحصائية لا تساعدنا في ذلك. لقد أكّد غير واحد بأنّ الباحثين غالباً ما يخطئون في تفسير نتائج اختبارات الدلالة الإحصائية، حيث يجعلون الحصول على النتيجة الدالة إحصائياً دليلاً على تأثير المعالجة (أو تأثير المتغيّرات المستقلة في المتغيّرات التابعة) . وقد أكد ذلك هوستون Huston(1993) حيث قال "الكثير من الباحثين والمحرّرين وطلاب الدراسات العليا يخلطون بين الدلالة الإحصائية لنتائجهم والدلالة العملية، ولذا فإنّهم يخفقون في التفكير الحذر في بياناتهم، وعلاقتها بالعالم الكبير والحقيقي"(p.5) .

وقد وضّح هذا الخطأ شيفر Shaver(1992) بقوله "من الخطأ تماماً استنتاج أنّ النتيجة الدالة إحصائياً تشير إلى تأثير المعالجة كما يشاهد غالباً في عبارات مثل الفرق الدال إحصائياً بين المتوسطات يشير إلى أنّ المعالجة (A) أكثر فعالية من المعالجة (B) ،مثل هذا الاستنتاج الخاطئ يُكرّس بواسطة كتب الإحصاء والتي تفترض أنّ اختبارات الدلالة الإحصائية تعبّر عن تساؤلات مثل (هل المعالجة فعّالة؟ هل العقار (X) يزيد النشاط أكثر من العلاج البديل " (p.15).

إنّ عبارة (تأثير) كما هي مستخدمة أحياناً في الإحصاء الاستدلالي مثل (التأثير الرئيسي، وجود تأثير، التأثير ليس معدوماً،...) غامضة وتتضمن السببية التي لا يمكن الانتباه لها إلاّ من خلال تصميم الدراسة وليس من خلال الاختبار الإحصائي، وكما قال كوهين Cohen : "التضمين السببي في عبارة تأثير يمكن أن يحجب الحقيقة بأنّ التفسيرات السببية لا ينتبه لها أبداً من خلال النتائج الإحصائية ولكنها تتطلب أسس منطقية وجوهرية " (Shaver,1992,p.10).

ولقد أفادت الدراسات الامبريقية للممارسة البحثية إلى أنّ الفهم السطحي لاختبار الدلالة الإحصائية قاد فعلياً إلى تشويه خطير مثل أن يفسّر الباحثون النتائج الدالة على أنّها ذات حجم تأثير عالي (Huston,1993). ومن سوء الحظ، أنّ ذلك تزامن مع نتائج لا يمكن محوها من تلك الممارسات التحريرية. الأخطاء من النوع (I) (رفض الفرضية الصفرية عندما تكون فعلياً صحيحة) يمكن أن تحدث نظرياً في واحدة من كل (٢٠) دراسة، باستخدام منطق ومستويات اختبار الدلالة الإحصائية الشائعة. عندما يعمل الخطأ من النوع (I)، توجد دلالة [إحصائية]، وبوضوح يشجع نشر الدراسة (Keaster,1988). وقد بدأ تايلور منذ عام ١٩٣١ بملاحظة النزعة نحو سوء تفسير الدلالة الإحصائية وقال: "التفسيرات التي عادة ما

تستنتج من الدراسات الحديثة تشير بوضوح إلى ميلنا للاعتقاد والتصوّر بأنّ الدلالة الإحصائية تكافئ الدلالة الاجتماعية، هذا المصطلحان مختلفان بالضرورة ويجب ألاّ يخلطا. الفروق التي هي دالة إحصائياً ليست دائماً دالة اجتماعياً، والفروق التي لم تلاحظ بأنّها دالة إحصائيا ربّما تكون دالة اجتماعياً" (Daniel,1997,p.12). ويرى دانييل Daniel(1997) أنّ الدلالة الإحصائية والتأثير شيئان مختلفان تماماً، وأنّ مصطلح (دال) خيار مَرَضي لأنّ القيمة الصغيرة لـ (r) لا تعني بالضرورة قوة العلاقة بين المتغيّرات المستقلة والتابعة في الدراسة.

وقد أوضح سنيدر ولاوسون Snyder and Lawson(1992) ذلك بقولهما "حدوث النتيجة الدالة إحصائياً لا تعني شيئاً حول قوة التلازم (Association) بين المعالجة والدرجات، النتيجة الدالة إحصائياً تقود إلى الاستدلال بأنّ بعض التلازم موجود، ولكن لا تعني أبداً بأي منطق وجود درجة مهمة من التلازم الجوهري. في المقابل، إثبات تلازم إحصائي قوي يمكن أن يحدث حتى لو كانت النتائج غير دالة إحصائياً، لعبة الاستدلال على الدرجة الحقيقية للتلازم الإحصائي نكتة، إنّه حجم العينة. لقد حان الوقت لتعريف غموض قوّة التلازم الإحصائي بوضوح أكثر، وربط هذه الفكرة بالفروق الحقيقية بين متوسطات المجتمع" (p.13).

وكما قال ماكلين وإرنست McLean and Ernest(1998) "الاعتراض الأهم الذي يركز عليه في أزمة الممارسات هو أنّ الباحثين يجب أن يقوّموا الأهمية العملية للنتائج وليس فقط الدلالة الإحصائية" (p.9). الدلالة الإحصائية تهتم بما إذا كانت النتيجة بسبب الصدفة أو تقلّبات المعاينة أمّا الدلالة العملية (وتسمى أيضاً الدلالة النظرية، أو الدلالة العلمية، أو الدلالة التربوية، أو الدلالة الاجتماعية، أو الدلالة الطبية، ...وهكذا) هي التي تهتم بما إذا كانت النتيجة مفيدة في عالم الواقع (Kirk,1996). ولذا فإنّ الفروق الحقيقية حتى لو كانت ضئيلة يمكن أن يحكم بأنّها دالة إحصائياً إذا كان خطأ المعاينة صغيراً بسبب حجم العينة الكبير، وعلى أي حال فإنّ اختبارات الفرضية الصفرية لا تخبرنا أي جزء من الفروق الدالة إحصائياً بسبب حجم التأثير وأي جزء بسبب خطأ المعاينة (Nix and Barnette, 1998). وكما قال كيلو Kellow(1998) وهو أحد العاملين في مجال التقويم: "لسوء الحظ، الكثير من البرامج لا تزال مستمرّة اعتماداً على حجم العينة المستخدمة لتقويم البرنامج وليس على الفعالية الحقيقية للبرنامج، يحدث هذا بسبب أنّ المقوّمين ليسوا مهتمين ولا منتبهين إلى أنّ اختبارات الدلالة الإحصائية التي يعتمدون عليها في مثل تلك التوصيات هي من صنيع حجم العينة الكبير بينما الفروق الفعلية لحجم التأثير ليس لها تأثير في الواقع الفعلي" (p.13).

ثالثاً : استخدام قيمة (r) لتقدير حجم التأثير

تعتبر هذه المشكلة هي مرحلة متقدّمة من المشكلتين السابقتين. في المشكلة الأولى ينظر الباحث إلى النتائج الدالة إحصائياً على أنّها مهمّة سواء في ذاتها أو بمقارنتها مع النتائج الأخرى غير الدالة إحصائياً. وفي المشكلة الثانية، ينظر الباحث إلى أنّ وجود دلالة إحصائية يعني وجود تأثيرات للمعالجة أو تأثيرات للمتغيرات المستقلة على المتغيرات التابعة. أمّا في هذه المرحلة المتقدمة من المشكلة يستعمل الباحث قيم احتمالية الدلالة الإحصائية المحسوبة ($r_{calculated}$) كمؤشر لمقدار التأثير.

يبدو أنّ الوصول السهل لقيم (r) عن طريق برامج الحاسب الإحصائية قاد إلى سوء فهم واستخدام هذه المعلومات، ولأنّ العديد من الباحثين يركزون بحثهم على قيم (r)، فإنّ خلطاً حول معنى قيمة (r) أصبح شائعاً في الأدبيات. وقد وضح ذلك كارفر وثومبسون بأنّ مستخدمي الاختبارالإحصائي للفرضية الصفرية غالباً ما يسيئون تفسير معنى قيمة (r) كما لو أنّها مقياس مقدار أو طول، ويظهر هذا من خلال عبارات مثل (غالباً تحقق الدلالة)، أو (عالية الدلالة) (Nix and Barnette,1998). ويرى شيفر Shaver(1992) بأنّ اعتماد قيمة الاحتمالية الإحصائية على حجم العينة يجعلها مؤشراً غير مناسب لمقدار النتيجة.

ويرى كيستر Keaster(1988) بأنّ اختبارات الدلالة الإحصائية وضعت للوصول لقرارات (نعم/لا)، ولذا يجب الامتناع عن محاولة تفسير مقدار نتيجة احتماليتها المحسوبة. وقد اعتبر لاي أنّ استعمال الباحثين مستويات (r) لمقارنة نتائج عدّة دراسات مشكلة توضح سوء استخدام اختبارات الدلالة الإحصائية، حيث يرى سنيدر ولاوسون أنّ القيمة الصغيرة لـ (r) لا تعني بالضرورة أنّ قوة العلاقة بين المتغيّرات المستقلة والمتغيّرات التابعة كبيرة في الدراسة (Snyder and Lawson,1992).

رابعاً – استخدام قيمة (r) كتفسير لاحتمالية الفرضية الصفرية نفسها

نتيجة اختبار الدلالة الإحصائية تخص احتمالية النتيجة نفسها وذلك إذا علم أنّ الفرضية الصفرية صحيحة، وهي بذلك ليست احتمالية أن تكون الفرضية الصفرية صحيحة إذا علمت تلك النتيجة المعيّنة (Shaver,1992).

فمثلاً إذا كانت قيمة الاحتمالية المحسوبة ($r_{calculated}$) هي ٠،٠٠٥ فهذا لايعني أنّ احتمالية أن تكون الفرضية الصفرية H_o صحيحة هي ٠،٠٠٥ ، إنّما يعني أنّ احتمالية الحصول على بيانات مختلفة عن الفرضية الصفرية H_o بنفس هذا المقدار أو أكثر من التي تمّ الحصول عليها في التجربة هي ٠،٠٠٥ ، وتكون الخطوة الاستدلالية من تلك النتيجة

($r_{calculated}=0.005$) لاستنتاج أنّ الفرضية الصفرية H_o خاطئة كما يلي: أنّ احتمالية أن تعطي بيانات هذه التجربة نتائج مختلفة بنفس المقدار أو أكثر عمّا هو في الفرضية الصفرية نادرة جداً (0.005)، طبعاً بافتراض صحة الفرضية الصفرية H_o ، وهـذا يعـني :

- وقوع حدث غير محتمل جداً.
- أنّ الفرضية الصفرية H_o خاطئة.

مجرّد الاستدلال بأنّ الفرضية الصفرية خاطئة يوضح أنّ هذه الاحتمالية ليست احتمالية الفرضية الصفرية H_o نفسها (Lane,2000)، (Cohen,1990). ولقد عرض لان Lane(2000) مثالاً عملياً وتفصيلياً يوضح ما سبق على النحو التالي:

لو أنّ إنساناً ادعى أنّ له قدرة على التنبؤ بجهة العملة عند رميها، وأنّ أحد المعترضين قاوم بشدّة هذا الادعاء وأراد أن يثبت بطلانه من خلال تجربة، وعليه فقد وضع الشخص المعترض فرضية صفرية هي أنّ ذلك الشخص المدّعي يستطيع أن يتنبأ بالوضع الصحيح للعملة في نصف المحاولات (H_o : $p = 0.5$). قام الشخص المعترض برمي العملة (٢٠) مرّة، واستطاع الشخص المدّعي أن يتنبأ وبشكل صحيح في (١١) مرّة. إذا كان ذلك الشخص المدّعي ليست لديه أي قدرة غير عادية واستطاع أن يتنبأ بشكل صحيح (١١) مرّة من (٢٠) مرّة، أي باحتمالية قيمتها ٠,٤١ ، فهل المعترض والذي عارض بشدّة ادّعاء ذلك الرجل يعتقد الآن أنّ احتمالية أن تكون الفرضية الصفرية صحيحة هي (٤١٪)؟ بالتأكيد لا، لن يصدّق المعترض بذلك، لأنّ الاحتمالية الأصلية لدى ذلك الشخص المعترض، والتي يعتقد فيها صحة الفرضية الصفرية H_o احتمالية عالية جداً لأن تكون صحيحـة (ربما تصـل إلى ٠,٩٩٩٩)، وليس لدى ذلك الإنسان المعترض استعداد للشك في اعتقاده في صحة فرضيته الصفرية H_o لمجرّد أنّ النتيجة التي حدثت لم تتفق كلياً مع فرضيته الصفرية H_o . إذاً التفسير المحتمل لما حدث هو أنّ الاختبار زوّدنا ببيانات تتفق مع الفرضية الصفرية H_o ولا تتعارض معها أو تنقضها، وهي أنّ ذلك الشخص ليست لديه قدرة خاصة، لأنّ الشخص الذي ليست له قدرة خاصة يمكنه أن يتنبأ بـ ٤٠٪ أو أكثر من المرّات وبشكل صحيح، ولذا لا يوجد سبب مشجع لهذا الشخص المعارض للاعتقاد بصحة ادعاء ذلك الرجل، وأنّ التجربة ببساطة تخفق في تزويدنا بإثبات على أنّ ذلك الشخص له قدرات خاصة تمكنّه من التنبؤ بشكل صحيح. إذاً الاحتمالية المحسوبة ($r_{calculated}$) هي احتمالية الحصول على بيانات مختلفة بمقدار معيّن أو بمقدار أكثر اختلافاً عن البيانات التي تم الحصول عليها فعلاً، أي هي احتمالية النتيجة نفسها عند صحة الفرضية الصفرية، أمّا الاحتمالية أن تكون الفرضية الصفرية H_o

صحيحة فلا تحدّد بالاختبار الإحصائي المنفّذ كجزء من اختبار الفرضيات (Lane,2000).

خامساً : استخدام قيمة (r) لتقويم إعادة النتيجة Replicability

يرى كيرك Kirk(1996) بأنّ هناك اعتقاداً واسعاً وغير صحيح يذهب إلى أنّ قيمة (r) هي احتمالية أن تكون النتيجة الدالة إحصائياً سوف توجد في إعادة التجربة أو الدراسة. ويرى نيكس وبارنت Nix and Barnette(1998) بأنّ بعض الكتب تشير إلى نفس الخطأ بقولها أنّ قيمة (r) للدلالة الإحصائية تشير إلى احتمالية أن تكون النتائج قابلة للإعادة، ويثبت هذا من خلال عبارات مثل (فروق ثابتة) أو(النتائج كانت ثابتة) علماً بأنّ هذا ليس من منطق اختبارات الفرضيات الصفرية، وأنّ قيمة (r) لا تخبرنا بأي شيء حول احتمالية إعادة الدراسة.

ويرى ثومبسون Thompson(1994) بأنّ اختبار الدلالة الإحصائية يتطلب استثمار جهد يفتقر إلى العائد المكافئ منه، وأنّ العلم هو مهمة عزل العلاقات التي تتكرر تحت ظروف محدّدة، بحيث تكون تلك المعرفة منشأة ويمكن أن تتراكم، إلاّ أنّ الدلالة الإحصائية لا تخبر بشكل كافٍ وصحيح ما إذا كانت النتائج في دراسة معطاة سوف تتكرّر، كعلماء ينبغي أن نسأل (أ) ما مقدار تأثيرات العينة (ب) وهل هذه النتائج سوف تعمّم، اختبار الدلالة الإحصائية لا يجيب عن أي من هذين التساؤلين، ولذا فإنّ الدلالة الإحصائية ربّما تحرف اهتمامنا عن الاعتبارات الأكثر أهمية.

وقد ذكر ماكلين وإرنست McLean and Ernest(1998) بأنّ جميع الأدبيات اعتبرت أنّ الإعادة عملية يخطئ الباحثون فهمها من خلال اختبارات الدلالة الإحصائية، وأنّ الدلالة الإحصائية تخبرنا بشيء ضار جداً في هذا الاتجاه. لكنّ كارفر Carver(1978) كان أكثر وضوحاً بقوله أنّه لا اختبارات الدلالة الإحصائية، ولا حجوم التأثير تزود الباحث باحتمالية إعادة النتائج في البحث المستقبلي.

سادساً : التحيّز للنتائج الدالة إحصائياً

عندما شاع في البحث التربوي اقتران الدلالة الإحصائية بالأهمية، وترسّخ الاعتماد الكلي على اختبارات الدلالة الإحصائية في تفسير النتائج واتخاذ القرارات وتحديد حجم تأثير المعالجات، ظهرت مشكلة أخرى جديدة وهي التحيّز لصالح الدراسات والأبحاث والمقالات العلمية التي تتضمن نتائج ذات دلالة إحصائية ضد الدراسات التي لم تصل إلى الدلالة الإحصائية. وقد لاحظ هذه المشكلة العديد من الباحثين، فقد ذكر ويلكنسون Wilkinson(1992) بأنّ تحيّز محرّري المجلات والمحكمين لنشر نتائج الأبحاث التي

وصلت للدلالة الإحصائية يجعلهم يعزّزون أسطورة الأهمية الكلية للدلالة الإحصائية في البحث، وأصبح غالباً ما تشاهد الاحتمالية أو كما يدعى (الدلالة) كمنتج نهائي للتحليل وليس الخطوة الأولى في شرح تلازم التغاير في البيانات. ولقد عبّر كيستر Keaster(1988) عن ذلك بقوله "الدراسات التي لها دلالة عالية لها فرصة كبيرة للظهور والتوثيق. ولقد بلغ الحد في التحيّز للدلالة الإحصائية -مع سوء فهم معناها - أنّ أحد المحررين يرفض نشر الأبحاث الدالة إحصائياً عند مستوى ٠.٠٥ لأنّه يفضل عوضاً عن ذلك النتائج التي تصل الدلالة الإحصائية فيها إلى مستوى ٠.٠١"(p.13). وقد نقل دانييل Daniel(1997) عن كامبل قوله "من الصعب انتشال المؤلفين بعيداً عن قيم (r) التي لديهم، فكلّما زادت الأصفار بعد الفاصلة العشرية، تشبث بهم المحرّرون"(p.13). وقد أدّى ذلك إلى أن يحاول الباحثون الحصول على نتائج ذات قيمة دلالة إحصائية منخفضة جداً، ليشير إليها بـ (عالية الدلالة) وذلك على حد قول دانييل Daniel(1997) ربما ليترك انطباعاً أفضل لدى محرّري المجلات والقراء. وتمضي هذه الممارسة إلى ما لانهاية لوضع أصفار أكثر على يمين الفاصلة العشرية لمحاولة جعل قيمة ($r_{calculated}$) أكثر جدارة بالملاحظة وبدون أي عمل مع الدلالة العملية للنتائج (Daniel,1997). وقد تهكم بوبهام على هـذه الممارسات بقوله أن بعض الباحثين يشيرون إلى احتمالياتهم كما لو كانت لوحة أهداف للعبة البيسبول ، مثل (r <0.000000001) (Daniel,1997). وبسبب هذا التحيّز للدلالة الإحصائية لم يعد من الغريب أن تجد في أدبيات الدراسات التي تقرّر النتائج غير الدالة إحصائياً عبارات مثل (تقترب من الدلالة)، وقد حمل هذا التحيّز من قبل الناشرين إلى أنّ العديد من الباحثين أصبحوا يبدون ولاءهم لاختبارات الدلالة الإحصائية على الملأ بينما هم يعرفون قصورها في أجوائهم الخاصة، كما أنّهم أصبحوا يشكلون دراساتهم ومخطوطاتهم حسب الطرق المقبولة في التحليل والتفسير بحيث تكون نتائجهم مصاغة بشكل قابل للنشر (Daniel,1997).

ولقد لاحظ هذا التحيّز هوستون Huston(1993) أيضاً وقال بأنّ الأدبيات النفسية منحازة ضد المقالات التي تخفق في وصف النتائج الدالة إحصائياً، وكذلك القرّاء والمحرّرون والمؤلفون، كلّهم يفضلون النتائج الدالة إحصائياً. ونتيجة لتلك الاتجاهات والممارسات، فإنّ الباحثين يمتنعون عن إرسال مقالاتهم التي لا تستطيع وصف دلالة إحصائية، و يخافون من الدفاع عن حقهم في أسطر لتوضيح رأيهم في قضية عامة اعتماداً على تلك النتائج التي ليست بها دلالة إحصائية، وذلك لأنّ محرّري المجلات النفسية يكرسون نفس الاتجاه بالنسبة للمقالات التي تصف والتي لا تصف نتائج دالة إحصائياً (Keaster,1988).

وكما قال بيلار Bailar بأنّ التفريق بين الدلالة الإحصائية والأهمية العملية يزداد أهمية عندما يكون هناك عجز في الوصول إلى الدلالة الإحصائية والذي ربما يساء تفسيره

بحيث يشار إلى تلك النتائج على أنّها قليلة الأهمية، ممّا يعني الحاجة إلى مزيد من التحاليل لتحديد الأهمية العملية (Slack,1997).

وقد لاحظ نيكس وبارنيت (Nix and Barnette(1998 أنّ نتيجة تلك الممارسات هي أنّه بدلاً من أن تسهم كل المحاولات البحثية في بناء المعرفة، فإنّ الدراسات المحظوظة فقط والتي وصلت إلى الدلالة الإحصائية عن طريق العينة كبيرة الحجم أو عن طريق الصدفة هي التي تصل إلى مجتمع الباحثين وأدبيات البحث، بينما يفترض أن تصل كل الدراسات التي صمّمت بشكل صحيح وتمّ ضبطها وقياسها إلى النشر وإلى أدبيات البحث بغض النظر عن دلالتها الإحصائية.

سابعاً : مشكلة صياغة التساؤلات البحثية والفرضيات الصفرية

فكرة اختبار الفرضيات تنص على أن يصيغ الباحث الفرضية الصفرية (مثل لا توجد فروق ذات دلالة إحصائية)، ثم يلاحظ إذا كانت البيانات التي حصل عليها تسمح له برفض الفرضية الصفرية أم لا، وكما يرى لان (Lane(2000 أنّ المشكلة تكمن أنّه في الطبيعة نادراً ما تكون فروق التأثيرات صفراً، ولذا فإنّ الشيء المهم هو تقدير مقدار التأثير وليس اختبار ما إذا كانت التأثيرات صفراً أم لا.

ويرى كيرك (Kirk(1996 بأنّ رفض الفرضية الصفرية لا يزودنا بالكثير من المعلومات، حيث أنّنا نعرف مسبقاً أنّ الفرضية الصفرية خاطئة، وعلى الرغم من ذلك نحن نحسب إحصاءة الاختبار الذي يمكننا من تحديد احتمالية الحصول على الفرق بنفس القدر أو أكبر من المشاهد، إذا كانت الفرضية الصفرية صحيحة. ويرى لاي بأنّ الباحث يستطيع صياغة الفرضية الصفرية H_o بشكل أكثر ملاءمة، فبدلاً من السؤال هل يوجد فرق (مجرّد فرق)، يفترض أن يسأل الباحثون عما إذا كانت هناك فروق عملية (تربوية)، وبدلاً من السؤال ما إذا كانت (نيسان Nissan) أفضل استهلاكاً للوقود من (كاديلاك Cadillac)، يفترض أن نسأل كم عدد الجوالين التي توفرها (نيسان) مقارنة بـ (كاديلاك Cadillac)، وهل هذا الفرق مهم عملياً؟ وبنفس الشكل بدلاً من السؤال ما إذا كانت إحدى المجموعات لها درجات أعلى من الأخرى، يفترض أن نسأل: ما مقدار فرق الدرجات التي حصلت عليها المجموعة العالية مقارنة بالمجموعة الأخرى؟ وهل هذا الفرق ذا أهمية عملية أو تربوية، إنّ صياغة الفرضية الصفرية إلى حد ما غير مناسبة لأنّها لا تبحث عن الأهمية، كما أنّ جعل الطلاب يختبرون الفرضيات يحرف انتباههم عن مقدار النتيجة إلى مقدار قيمة (r)، الشيء الأكثر أهمية هو تقدير مقدار التأثير وليس اختبار ما إذا كان التأثير صفراً (Lane,2000).

ويؤكد سنايدر وثومبسون Snyder and Thomson(1998) أنّ إحدى مشكلات الدلالة الإحصائية هو التوجه الغالب لاستخدام الفرضية الصفرية لعدم وجود فروق أو العلاقة الصفرية، وأنّه عندما تستخدم هذه الفرضية، ويفترض أن تكون تأثيرات المجتمع صفراً، فإنّ الكثير من الإحصائيين يعتقدون بأنّ الفرضية الصفرية لعدم وجود فروق مستحيل أن تكون صحيحة في المجتمع، وأنّه دائماً ما تكون هناك بعض الفروق في معالم المجتمع حتى لو كانت تلك الفروق ضئيلة. وقد قال هايز: "بالتأكيد، لا يوجد شيء على الأرض مستقل بشكل تام عن أي شيء آخر، قوة التلازم ربما تقترب من الصفر، ولكن نادراً أو مستحيل أن تكون صفراً تماما، وهذا يعني بأنّ التأثيرات غير الصفرية للعينة دائماً متوقعة، وأنّ أي دراسة ربما تصل إلى نتائج دالة إحصائياً إذا استخدم الباحث عناصر كافية (Snyder and Thomson,1998).

وقد حاول ناب Knapp(1998) أن يوضح بأنّ الفرضية الصفرية هي الفرضية التي تُختبر، وأنّها لا يشترط أن تكون صفراً أو تتضمن الصفر في أي مكان، لكنّ البعض من الباحثين أمثال نيكس وبارنيت Nix and Barnette(1998) علّق على مثل ذلك الرأي بقولهما أنّ الفرضية طبعاً هي الفرضية التي تُختبر، ولكن أن نقول بأنّ الفرضية الصفرية لا يجب أن تكون صفراً فإنّ هذا لغز.

ويرى كيرك Kirk(1996) بأنّ هناك انتقاداً رئيساً آخراً لاختبار الفرضية الصفرية التقليدية وهو أنّ الإجراء لا يخبر الباحثين بما يريدون أن يعرفوه، بمعنى أنّ اختبار دلالة الفرضية الصفرية والإحصاء العلمي يوجه تساؤلات مختلفة، ففي الإحصاء العلمي ما نريد أن نعرفه هو احتمالية أن تكون الفرضية الصفرية H_o صحيحة عندما تكون لدينا مجموعة من البيانات D، أي $p(H_O|D)$، بينما ما يخبرنا به اختبار الفرضية الصفرية هو احتمالية الحصول على تلك البيانات -أو بيانات أكثر اختلافاً- إذا كانت الفرضية الصفرية صحيحة $p(D|H_O)$. ويرى كيرك بأنّه لسوء حظ الباحثين فإنّ البيانات المحصلة والتي تكون فيها $p(D|H_O)$ منخفضة لا تستلزم أن تكون $p(H_O|D)$ أيضاً منخفضة، وأنّ الباحثين يستنبطون وبشكل خاطئ أنّه إذا كانت قيمة r لاختبار إحصاءة منخفضة مثل أقل من ٠,٠٥ فإنّ الفرضية الصفرية ربما تكون خاطئة، وقال كيرك بأنّ هذا النوع من الاستنتاج أشار إليه جرينبوم وفالك وسمياه خداع الإثبات الإحصائي المتناقض (Kirk,1996).

ثامناً: مشكلة الطبيعة الثنائية لنتائج اختبارات الدلالة الإحصائية

يرى هوستون Huston(1993) بأنّ الاعتماد العام على قرار اختبار الدلالة الإحصائية، يقود إلى الانطباع بأنّ هناك خطأً حاداً موجوداً بين الفروق الدالة إحصائياً وغير الدالة إحصائياً. ولقد أعتبر كيرك Kirk(1996) بأنّ من النقد لاختبار الدلالة الإحصائية هو أنّه بتبنّي دلالة ذات مستوى ثابت فإنّ الباحث يحوّل قيمة الاحتمالية المتصلة Continuum uncertainty إلى قيمة ثنائية(رفض أو عدم رفض القرار)، وأنّ استخدام هذه الاستراتيجية للقرار يمكن أن يقود إلى وضع شاذ يكون فيه باحثان يحصلان على تأثيرات معالجة متماثلة، ولكن يستخلصون استنتاجات مختلفة من بحثهما. على سبيل المثال قد يحصل أحد الباحثين على قيمة r =0.06 ، ويقرّر عدم رفض الفرضية الصفرية، بينما يحصل الباحث الآخر الذي يستعمل عينات أكثر بقليل على قيمة r =0.05 والتي تقود إلى رفض الفرضية الصفرية، المشكلة هنا هو أنّ تأثيرات متماثلة للمعالجة تقود إلى قرارات مختلفة (Kirk,1996). وكما قال نيكس وبارنيت Nix and Barnette(1998) بأنّ التبسيط في قانون القرار الثنائي المحدّد بقيم (r) مغرٍ، ولكنه يمكن أن يقود إلى خطأ في تفسير الدلالة الإحصائية والأكثر أهمية هو أنّه قد يصرفنا عن الهدف الأعلى للتساؤل، وهو تحديد هل لنتائج الاختبار أي قيمة عملية أم لا. وقد نبّه روسناو وروسنتال Rosnow & Rosenthal(1988) إلى خطورة التركيز على التفكير الثنائي (نعم / لا) في اختبار الدلالة الإحصائية، وقد أشار إلى ذلك هوبرتي حيث عارض استعمال الدلالة الإحصائية للوصول إلى إجابة محدّدة، وبدلاً من ذلك اقترح أن تستخدم اختبارات الدلالة الإحصائية كمرشد في تفسير النتائج (Wilkinson,1992).

إجابة السؤال الثاني

تمهيد

قبل الشروع في الإجابة عن السؤال الثاني وتحديد المفاهيم الإحصائية التي يمكن أن تقدّم حلولاً لأهم مشكلات الدلالة الإحصائية في البحث التربوي التي سبق تحديدها في إجابة التساؤل الأول من هذه الدراسة، هناك موقفان لعلماء الإحصاء المعاصرين تجاه اختبارات الدلالة الإحصائية، سيعرضهما الباحث بين يدي إجابته على التساؤل الثاني على النحو التالي:

الموقف من اختبارات الدلالة الإحصائية

أولاً – الاتجاه المعارض

يرى بعض الباحثين المهتمين بمشكلات اختبارات الدلالة الإحصائية في البحث التربوي بضرورة إعادة النظر في التقليد العلمي المتبّع حالياً في البحث التربوي الذي يركز على اختبار الفرضيات من خلال استخدام اختبارات الدلالة الإحصائية. وقد عبّر عن مثل هذا الرأي كيستر Keaster(1988) حيث ذكر أنّ علماء الاجتماع يجب أن يعملوا بشكل أفضل لإعادة فحص أغراضهم من استخدام اختبارات الدلالة الإحصائية، وأن يحاولوا أن يستنبطوا طرقاً أفضل للوصول إلى تلك الأغراض بدلاً من الاستمرار في اللجوء إلى تكنيكات التي هي في أحسن أحوالها تضلّل البحث الامبريقي.

وقد ذهب كارفر Carver(1993) إلى حد القول بأنّ عدم استعمال اختبارات الدلالة الإحصائية سيكون أفضل للبحث التربوي حيث ذكر بأنّه إذا استطاع أحد ألاّ يستخدم الدلالة الإحصائية مرّة أخرى لتحديد دلالة الفرق، فإنّ الباحثين سيضطرون لاستخدام التصميمات التي تظهر بوضوح الأهمية العلمية للفروق، ويرى بأنّه بدون استخدام الدلالة الإحصائية، يضطر الباحثون إلى تلمس مشكلات الاستدلال العلمي بدلاً من الالتزام باختبار الدلالة الإحصائية، وأنّ البحث التربوي سوف يكون أفضل إذا أوقف اختبار الدلالة الإحصائية لنتائجه. وقد ذهب كرونباخ إلى القول بأنّ الوقت قد حان للتخلي عن الفرضية الصفرية، بينما وصف شلمان حال الباحثين التربويين في اعتمادهم على اختبارات الدلالة الإحصائية بقوله إنّ الباحثين التربويين أسلموا أنفسهم لعبودية اختبار الفرضية الإحصائية (Keaster,1988).

لقد دعت تلك النظرة لاختبارات الدلالة الإحصائية العديد من الباحثين إلى القول بخطر اختبارات الدلالة الإحصائية والدعوة لحظرها ومنعها من النشر، ومن أمثال هؤلاء

كارفر و ميهل وسميدث وروزبوم، ولقد عبّر روزبوم عن رأيه بقوله "اختبار الدلالة الإحصائية بالتأكيد هو أكبر إجراء مضلل وساحق للعظام يدرّس للطلاب، وإنّه من عجائب علم الاجتماع أن تبقى هذه الممارسة الإحصائية غير متأثّرة ولا مستجيبة للنقد" ، بينما قال سميدث: "اختبار الدلالة الإحصائية يعيق نمو المعرفة العلمية، وهي لا تقوم أبداً بأي إسهام إيجابي" (Keaster,1988,p.33).

وقد ذهب إلى هذا الاتجاه نيكس وبارنت Nix and Barnette(1998) حين قالا "نعتقد أنّ أغلب الإحصائيين يجب أن يرحبوا بالتغيير والذي سيقود إلى منع اختبار الفرضيات الصفرية... إنّ الأدبيات الإحصائية الحالية تشهد على الوعي المتنامي من أنّ فائدة طرق اختبار الفرضيات المعروفة قاصرة ومحدودة بشكل حاد... وإنّ نتيجة فقد اختبارات الفرضيات الصفرية الإحصائية لن يعيق نمو المعرفة العلمية، بل سيتقدّم بالمعرفة" (pp.12-13). ولقد أشار ويلكنسون Wilkinson(1992) إلى ضرورة التخلّي عن تقاليد اختبارات الدلالة الإحصائية بقوله "الباحثون يجب ويحذر أن يعتبروا ما تعنيه نتائج أبحاثهم في الحقيقة، إجراءات اختبارات الدلالة الإحصائية غالباً ما تستعمل كبديل للتفكير أكثر منها كمعاونة في عملية التفكير... إنّه الوقت للرجوع إلى المعنى الأساسي العام" (p.819).

ولعلّ الباحث يختم آراء هذا الاتجاه بواحدة من أكثر النظرات تشدداً في الموقف من اختبارات الدلالة الإحصائية والتي صورت بواسطة ميهل حيث قال "السيد رونالد Sir Ronald أربكنا ونومنا مغناطيسياً وقادنا بانحدار نحو العبث، أنا أعتقد أنّ الاعتماد الكلي على أسلوب رفض الفرضيات الصفرية كطريقة معيارية لتأييد النتائج الجوهرية في ميدان العلوم الإنسانية هي مشكلة فظيعة واستراتيجية علمية هزيلة وواحدة من أسوأ الأشياء التي حدثت في تاريخ علم النفس... لن أقوم بعمل تصحيح للإحصائيين، بل أقول إنّ كل العمل فيه خلل جوهري بحيث يعتبر علمياً عمل تافه وأحمق" (Wilkinson,1992,pp.817-823).

ثانياً – الاتجاه المدافع

جميع المدافعين عن الاستخدامات الحالية لاختبارات الدلالة الإحصائية لم يتمكنوا من تجنّب الاعتراف بقصور ومحدودية اختبارات الدلالة الإحصائية، وقد ركز المدافعون على فكرة أنّ اختبارات الدلالة الإحصائية ليس لديها مشكلة في ذاتها، ولكن المشكلة عندما تستخدم كمؤشرات للوصول إلى إجابات ومقارنات محدّدة، ولذا فهم يؤيدون استخدامها إذا كان بشكل صحيح ومع المعرفة بحدودها (Daniel,1998a).

وقد تمثّل هذا التوجه في تقبّل ودعم استعمال اختبار الدلالة الإحصائية كأحد المعايير التي يجب أن تتحقق حتى يثبت الموقف امبريقياً، بحيث تستعمل بشكل محدود ومدعومة بمؤشرات أخرى (McLean and Ernest,1998; Nix and Barnette,1998b; Knapp,1998; Daniel,1998b;Thompson,1998b).

ويرى نيكس وبارنت Nix and Barnette(1998) وكيرك Kirk(1996) بأنّ اختبارات الدلالة الإحصائية يجب أن ترفق بمقياس واحد على الأقل للدلالة العملية، وإذا كان ممكناً يفضل إثبات الإعادة الخارجية أيضاً.

لقد شبّه سون موقع اختبارات الدلالة الإحصائية بين الأساليب الإحصائية الأخرى التي يستعملها الباحثون في أبحاثهم بـ"الضيف المتغطرس" في حفلة عشاء والذي -على حد قوله - يستحوذ وبشكل غير مناسب على جميع الحديث والأنشطة إلى الحد الذي ننسى فيه من هو المضيف، نحن لا نستطيع ألاّ ندعو هذا الضيف، وبدلاً من ذلك نحن نحتاج إلى وضع هذا الضيف في المكان المناسب، وبالتحديد كواحد من الضيوف العديدين، ولن يتحول بأي حال إلى أن يكون هو المضيف"، وقد علّق ماكلين وإرنيست على تمثيل سون السابق بقولهما : "إشارة سون Suen إلى الوضع الصحيح والمناسب هي دعوة لاعتبار اختبار الدلالة الإحصائية وسيلة لفلترة تقلبات المعاينة، بحيث تكون المعلومات الملاحظة (سواء فروق أو ارتباطات) أكثر وضوحاً وتحديداً... الضيوف الآخرين الذين يجب على الباحث أن يرفعهم إلى مستوى أعلى يتضمنون: التأكد من نوعية تصميم البحث، وثبات المقياس، واستعمال مقاييس حجم التأثير " (McLean and Ernest,1998,p.17).

استنتاج الحلول البديلة

بناء على ما ناقشه الباحث في الإطار النظري عن مفهوم الدلالة الإحصائية والعوامل المؤثرة في نتائجها، وبناء على ما توصل إليه الباحث في إجابة التساؤل الأول من تحديد لأهم مشكلات الدلالة الإحصائية، وبعد عرض اتجاهات علماء الإحصاء تجاه اختبارات الدلالة الإحصائية، ونظر الباحث في الكثير من أدبيات البحث التربوي التي تناولت أطرافاً من مشكلة البحث، توصل الباحث إلى إجابة السؤال الثاني من خلال تحديد الحلول التالية:

- القوة الإحصائية.
- حدود الثقة.
- حجم التأثير.
- تحليل الإعادة.

وسيقوم الباحث باستعراض هذه الحلول البديلة ومناقشتها بشكل تفصيلي على النحو التالي:

الحلول البديلة

أولاً – القوة الإحصائية Statistical Power

مفهوم القوة الإحصائية

تحليل القوة الإحصائية هو أحد الاختبارات المكملة لاختبارات الدلالة الإحصائية، وهو يقدّر احتمالية الخطأ من النوع (II) (عدم رفض الفرضية الصفرية H_o عندما تكون خاطئة) (Wilkinson,1992). وقد عرّفها هوستون Huston(1993) بأنّها "الاحتمالية أن يرفض الاختبار الفرضية الصفرية بطريقة صحيحة"(p.13).

وقد لاحظ كوهين وشيفر أنّ هذا التوجه نادراً ما يستخدم من قبل الباحثين، وهو تحليل يتضمن التعامل مع a، وحجم التأثير، وحجم العينة n، و b (بيتا)، وقد لاحظ ثومبسون أنّ القوة إذا كانت عالية، فإنّ النتائج غير الدالة إحصائياً تبقى تسهم في أساس المعرفة، وكما يرى لاي بأنّ أي باحث يقبل الفرضية الصفرية بدون أن يعرف قوة الاختبار الإحصائي فإنّه يكون مسؤولاً عن حدوث نسبة خطأ كبيرة من النوع (II)، وأنّ الباحثين الذين يؤيدون وبطريقة غير صحيحة قبول الفرضية الصفرية إذا لم يلاحظوا أي دلالة إحصائية دون اعتبار للقوة الإحصائية يسيئون قيادة القرّاء (Wilkinson,1992). ويرى نيكس وبارنيت Nix and Barnette(1998) بأنّه إذا كان لدينا اختبار منخفض القوة فإنّه لن يمكننا تحسّس والتقاط الدلالة الإحصائية، بينما إذا كان لدينا اختبار عالي القوة فإنّ حجم الفروق بغض النظر عن صغرها سوف تكون دالة إحصائياً.

ويعتقد الكثيرون، أنّ التركيز على التحكم في الخطأ من النوع (I) المستعمل في الأساليب الشائعة، ربما يكون السبب وراء رؤية العديد من الدراسات الإحصائية التي تعاني من نقص في القوة، وعلى العموم فإنّ التحكم في الخطأ من النوع (I) يجعل الباحثين يدفعون الثمن وهو العجز في التحكم في الخطأ من النوع (II) أي النقص في قوة الاختبار، ممّا يعني الفرصة الأقل للحصول على نتائج دالة إحصائياً (Nix and Barnette,1998).

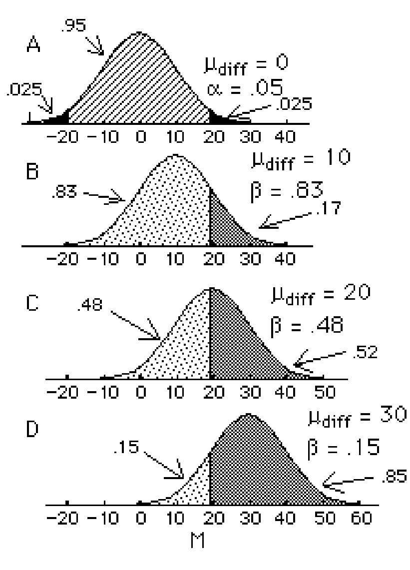
ولقد دلت دراسات لكوهين، وسيدليمر وقفرينيزر أنّ الباحثين لديهم فرصة أقل من ٥٠٪ لرفض الفرضية الصفرية، وهذه النتائج توضح التخريب للمعرفة التراكمية الذي يحدث في العلوم الاجتماعية، فإذا كان المؤلفون ليس لديهم إلاّ فرصة ٥٠٪ تقريباً لرفض الفرضية الصفرية والحصول على نتائج دالة إحصائياً، وإذا كانت هذه الدراسات ستعاني بشدة من النشر بسبب التحيّز الموجود لنشر النتائج الدالة إحصائياً، فإنّ هذا يعني وجود آلاف من

الدراسات التي يكون فيها حجم التأثير له معنى رفضت من النشر أو أنّها أصلاً لم ترسل للنشر وأنّ تلك الدراسات فقدت قيمتها لأنّها لم تتجاوز الدلالة الإحصائية، وهذا يعني أنّ البحث التربوي فقد الكثير من التحسينات والتطويرات في التجارب، علماً بأنّه يجب ملاحظة أنّ دراسة القوة مفهوم منطقي فقط في سياق اختبارات الدلالة الإحصائية، وأنه إذا لم تستخدم اختبارات الدلالة الإحصائية فإنّ مفهوم القوة الإحصائية ليس له مكان أو معنى (Nix and Barnette,1998). ويرى كل من ماكسويل وديلاني بأنّ إخفاق الدراسات في الأخذ بإيجابيات تحليل القوة يعتبر مشكلة أكثر من مجرد مشكلة تأثر الدلالة الإحصائية بحجم العينة (Huston,1993).

العوامل المؤثرة في القوة الإحصائية

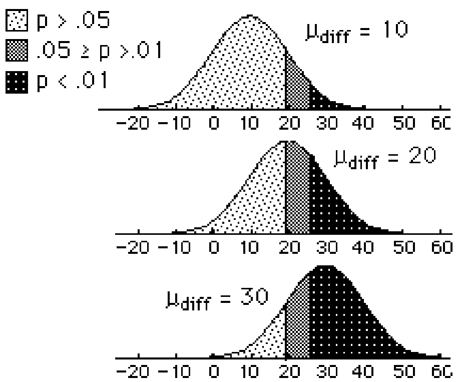
قوة الاختبار الإحصائي (أو احتمالية ملاحظة النتائج الدالة إحصائيا)، تعتمد على عدّة عوامل هي:

١ - حجم التأثير (أو مقدار الفرق بين المتوسطات): عندما يكون حجم التأثير المراد الكشف عنه كبيراً، فإنّ الاختبار الإحصائي يكون أكثر قوة ممّا لو كان الاختبار معد للكشف عن حجم التأثير الصغير جداً، الشكل (٢) يوضح قيمة القوة الإحصائية في أربع حالات: في الحالة A عندما تكون قيمة الفرق بين المتوسطات = صفر فإنّ قيمة القوة الإحصائية = ٠,٠٥، أمّا في الحالة B عندما تكون قيمة الفرق بين المتوسطات=١٠ فإنّ القوة الإحصائية = ٠,١٧ ، وهكذا حتى في الحالة D عندما كان مقدار الفرق بين المتوسطات=٣٠، أصبحت قيمة القوة الإحصائية ٠,٨٥ (Lane,2000). قد يعتقد البعض أنّ حجم التأثير يخرج عن تحكم الباحث، لكن في الواقع أنّ الباحث باختياره مستويات معيّنة للمتغير المستقل، ربّما يزيد من حجم التأثير ومن ثمّ القوة الإحصائية، فعلى سبيل المثال،



شكل ٢: أثر حجم التأثير على القوة الإحصائية

عندما يقارن الباحث جرعة ٥٠٠ ملجم مع الجرعة البديلة فإنّ حجم التأثير ربما يكون أكبر ممّا لو قارن الباحث جرعة ٣٠٠ ملجم مع الجرعة البديلة (Lane,2000).



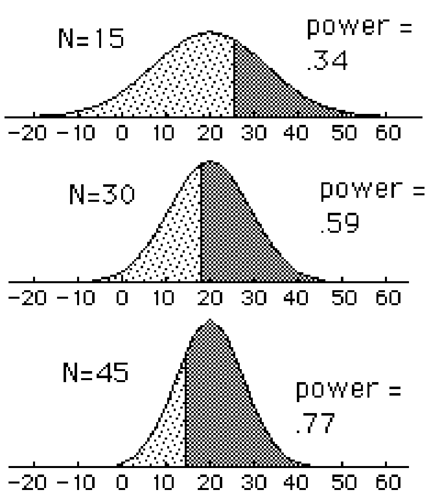
شكل ٣ : أثر مستويات الدلالة على القوة الإحصائية

٢ - مستوى الدلالة a : يتضح من الشكل (٣) أنّ قيمة القوة الإحصائية تنخفض كلما كانت مستويات الدلالة أكثر تحفظاً وانخفاضاً، كما يوضح الجدول (٩) أيضاً قيمة القوة الإحصائية عند تغيّر مستويات الدلالة(٠,٠٥ أو ٠,٠١) وكذلك عند تغيّر حجم التأثير (Lane,2000).

جدول ٩ :القوة الإحصائية عند مستويات مختلفة من الدلالة

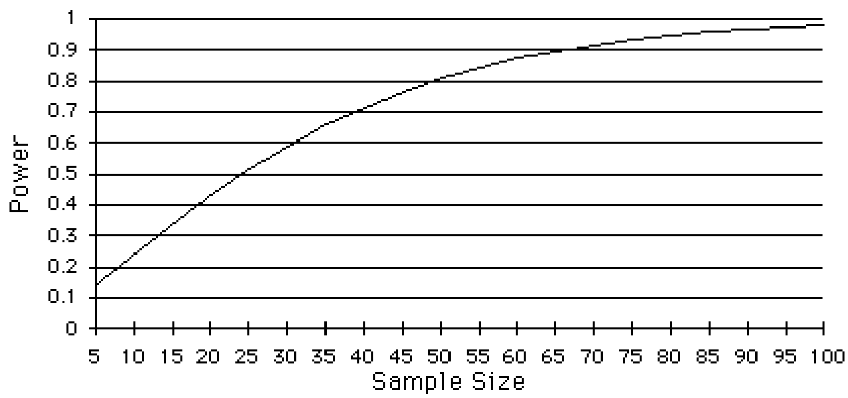
μ_{diff}	.05	.01
10	.17	.07
20	.52	.28
30	.85	.66

يتضح من الشكل (٣) والجدول (٩) أيضاً أنّ قيمة القوة الإحصائية ترتفع بزيادة مستوى الدلالة الإحصائية.



شكل ٤ : أثر حجم العينة على القوة الإحصائية

٣ - حجم العينة : الشكل (٤) يوضح العلاقة بين القوة الإحصائية وحجم العينة، حيث يتضح من الشكل أنّ القوة الإحصائية تزداد بازدياد حجم العينة، وهذا يعني أنّ القوة في دراسة ما قد تكون منخفضة بسبب انخفاض عدد أفراد عينة الدراسة، ويوضح الشكل (٥) أيضاً تلك العلاقة (Lane,2000).



شكل ٥ : العلاقة بين القوة الإحصائية وحجم العينة

٤ - مقدار التباين s^2 : تقل القوة الإحصائية بارتفاع مقدار التباين، فعلى سبيل المثال قوة الاختبار باستعمال مستوى دلالة (٠,٠٥) ، وحجم عينة (٢٥)، وفرق بين المتوسطات (٢٠)، عند قيم مختلفة من التباين موضحة بالجدول(١٠) (Lane,2000). هذا يعني أيضاً أنّ القوة الإحصائية تقل بانخفاض معامل الثبات (Nix and Barnette,1998) .

σ	Power
50	.52
75	.26
100	.17
125	.12
150	.10

جدول١٠: القوة الإحصائية عند قيم مختلفة من التباين

فوائد تحليل القوة الإحصائية

يرى كوهين Cohen بأنّ تحليل القوة الإحصائية يستطلع العلاقات بين المتغيرات الأربعة المطلوبة في الاستدلال الإحصائي وهي :

(١) حجم العينة n (٢) مستوى الدلالة a

(٣) حجم التأثير (٤) القوة الإحصائية.

وأنّه لأي أسلوب إحصائي فإنّ كل واحدة من هذه العلاقات تعتبر دالة للثلاث الأخرى (Huston,1993)

ووفقاً لما نقله هوستون Huston(1993) عن روساي Rossi فإنّ فوائد القوة هي:

١ - معرفة احتمالية الحصول على نتيجة دالة إحصائياً، ممّا يساعد الباحث في تحديد مدى الجدوى من تنفيذ الدراسة وذلك لتوفير الوقت والمال والجهد.

٢ - معرفة القوة يساعد في تفسير النتائج الصغيرة وشبه المعدومة (null results)، على سبيل المثال، إذا كانت القوة لدراسة ما صغيرة، فإنّه يمكن الاقتراح بأنّه لا توجد فرصة جيدة لرفض الفرضية الصفرية، ولذا فإنّ الاخفاق في رفض الفرضية الصفرية يجب ألاّ يقودنا سريعاً إلى الفرضية البديلة.

٣ - القوة تزودنا بمعلومات مفيدة حول البحث بشكل كلي، فعندما يكون متوسط القوة الإحصائية لكيان الأدبيات منخفضاً، فإنّ النتائج ستكون محل تساؤل.

وقد أضاف هوستون Huston(1993) إلى ما سبق أنّ تحليل القوة يجبر الشخص أن يأخذ في اعتباره حجم التأثير للدراسة، لأنّ حجم التأثير جزء أساسي من تحديد القوة ومن ثمّ زيادة الوعي بأهمية تقديرات حجم التأثير، كما أنّها تساعد الباحث ليكون أكثر إدراكاً للعلاقة بين العوامل الأربعة التي تحدّد القوة الإحصائية.

ويرى هوستون Huston(1993) نّ قصور تحليل القوة يكمن في عدم استخدامها من قبل كثير من الباحثين، ولعلّ سبب ذلك هو أنّه بمجرّد الحصول على نتيجة دالة إحصائياً يصبح من غير الممكن الحصول على الخطأ من النوع (II)، ومع التحيّز الموجود حالياً تجاه نتائج الدلالة الإحصائية والذي يمارس من قبل الغالبية العظمى لمجلات العلوم الاجتماعية، فإنّ وصف تحليل القوة يصبح ليس له علاقة بالموضوع. وقد أدى ذلك على حد تعبير روسنتال Rosenthal إلى ظهور مشكلة سماها (مشكلة درج الملفات The file drawer problem) حيث تكون المجلات ممتلئة بـ ٥٪ من الدراسات التي تتلبس بالخطأ من النوع (I)، بينما الأدراج الخلفية للملفات في المعمل أو المكتب ممتلئة بـ ٩٥٪ من الدراسات التي تتضمن عدم الدلالة الإحصائية (مثل نتائج $r > 0.05$) (Huston,1993).

ولتقدير القوة الإحصائية لاختبارات الدلالة الإحصائية الأكثر شيوعاً، يمكن الرجوع لجداول كوهين Cohen(1988) في كتابه الشهير (تحليل قوة الاختبار الإحصائي في العلوم السلوكية Statistical Power Analysis for the Behavioral Sciences)، كما توجد عدّة برامج حاسب آلي تساعد في حساب قيمة القوة الإحصائية، ومنها البرنامج المجاني الذي يقدمه قسم الإحصاء بجامعة UCLA بلوس أنجلوس بكاليفورنيا على إحدى صفحات موقعها على الإنترنت على العنوان التالي: http://www.stat.ucla.edu ،وذلك لتقدير القوة الإحصائية كما هو موضح بالملحق رقم(٤) .

ثانياً – حدود الثقة Confidence Interval

مفهوم حدود الثقة

عندما يبدأ الباحث بإحصاءة مشتقة من عينة فإنّ تلك الإحصاءة هي تقدير للقيمة في المجتمع، وإذا سحبت عينة مختلفة سوف يحصل على قيمة مختلفة. الطريقة الوحيدة للحصول على قيمة المجتمع هو قياس كل عنصر في المجتمع، وحتى لو كان هذا ممكناً، فإنّه سيكون هدراً للموارد (Lane,2000).

إذا كانت العينة صغيرة وذات قيم متفاوتة، فإنّ متوسط العينة ربما يكون بعيداً نسبياً عن متوسط المجتمع، أمّا إذا كانت العينة كبيرة وذات تفاوت أقل، فإنّ متوسط العينة ربما يكون قريباً جداً لمتوسط المجتمع (Hotulsky,1995). في حدود الثقة، يستخدم الباحث العينة لحساب المدى الذي يحتمل أن تقع فيه قيمة المجتمع وبمستوى معيّن من الثقة (عادة ما تكون ٩٥٪)، ويجب ملاحظة أنّ هناك قيمة حقيقية واحدة، وأنّ حدود الثقة تعرّف المدى الذي يمكن فيه للاحتمالية أن تحدث، وليست التغيّر في القيمة الحقيقية كما هو في الانحراف المعياري (Lane,2000).

القيمتان في كلا الحدين تسميان حدود الثقة، وكل القيم بين حدود الثقة تكوّن مدى الثقة الذي يتسع أو يضيق حسب مستوى الثقة المختارة، فعلى سبيل المثال عندما يكون مستوى الثقة ٩٥٪ فإنّ المدى يكون أوسع ممّا لو كان مستوى الثقة ٩٩٪، والعكس صحيح (Hotulsky,1995).

ويرى نيكس وبارنيت Nix and Barnette(1998) بأنّه إذا كانت اختبارات الدلالة الإحصائية تزودنا فقط بمعلومة حول ما إذا كانت الصدفة هي التفسير أو عدم التفسير للفروق الملاحظة، فإنّ استعمال حدود الثقة يعتبر بديلاً مناسباً لاختبارات الدلالة الإحصائية، والسبب أنّ كلا الطريقتين تزودان الباحث بنفس النتيجة (رفض أو عدم رفض الفرضية الصفرية)، علماً بأنّ حدود الثقة تزود الباحث بمعلومات إضافية لا تقدمها اختبارات الدلالة الإحصائية، ممّا يساعد الباحث في تفسير أفضل للنتائج، علماً بأنّ حدود الثقة لا تتطلب أي معلومات أخرى إضافة لما تتطلبه اختبارات الدلالة الإحصـائية.

ويرى كيرك Kirk(1996) بأنّه من المدهش وبالاعتماد على مراجعات الأدبيات الحديثة أنّ امتياز وفائدة طريقة حدود الثقة لم تلق اهتماماً من قبل مجتمع الباحثين، أي أنّه تم تجاهلها. ويرى نيكس وبارنيت Nix and Barnette(1998) بأنّه يجب على الباحث أن يصف في دراسته حدود الثقة بشكل روتيني لأنّها تخدم في تذكير الباحث بالأخطاء في

النتائج، ومدى الحاجة إلى تحسين القياس وطرق المعاينة، وأهم من ذلك أنّها تزود الباحث بالأساس لتعيين الحجم المناسب من حجم العينة، وهي أيضاً تماثل قوة الاختبار لأنّ الحجم الأكبر للعينة، والقوة الأعلى للاختبار سوف تكون فيها حدود الثقة أصغر، بينما حجم العينة الأصغر وقوة الاختبار الأقل سوف تكون حدود الثقة لهما أكبر.

وقد أوضح ويلكنسون Wilkinson(1992) بأنّ مستوى الثقة هي عبارة عن (واحد ناقص مستوى a)، فإذا مستوى الدلالة a موضوعة عند ٠,٠٥ فإنّ مستوى الثقة سوف تكون (١ -٠,٠٥) أي (٠,٩٥) ، وأنّ حدود الثقة تحسب حسب المعادلة العامة التالية:

حدود الثقة = الإحصاءة ± القيمة الحرجة × الخطأ المعياري

فإذا كان أحدهم يريد أن يقدّر متوسط المجتمع m بنسبة ثقة ٩٥٪، فإنّ الإحصاءة ستكون هي متوسط العينة x ، والخطأ المعياري $s_{\bar{x}}$ يمكن أن يحسب كالآتي:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

بينما تحدّد القيمة الحرجة حسب جدول التوزيع الطبيعي في كتب الإحصاء الذي يستخدم للحصول على القيم الحرجة المتعلقة بمستوى الثقة المرغوبة (Wilkinson,1992).

وللتوضيح بمثال، افترض أنّ المتوسط X يساوي ٣٠، والانحراف المعياري يساوي٢، وحجم العينة n هو ٤٠، والخطأ المعياري المحسوب $s_{\bar{x}}$ يساوي $\frac{2}{\sqrt{40}} = \frac{2}{6.324}$ = 0.32 ، فإنّ معادلة حدود الثقة عند مستوى ٩٥٪ هي:

حدود الثقة = $x \pm (1.96) s_{\bar{x}}$

= ٣٠ ± ١,٩٦ × ٠,٣٢

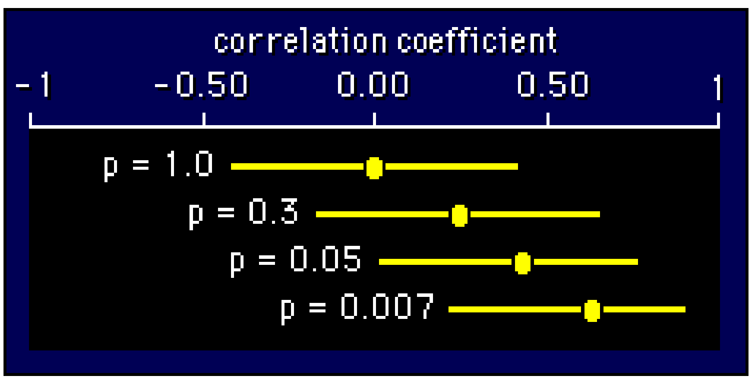
= ٣٠ ± ٠,٦٣

= ٢٩,٣٧أو ٣٠,٦٣

وهذا يعني أنّ الباحث يستنتج أنّه عند حدود ثقة ٩٥٪ ، فإنّ المدى من ٢٩,٣٧ إلى ٣٠,٦٣ يتضمن معلم المجتمع m .

تفسير الدلالة الإحصائية باستخدام حدود الثقة

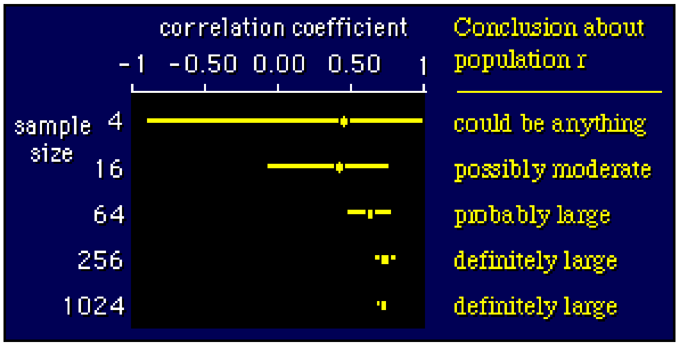
افترض أنّ الارتباط بين الطول والوزن لـ (٦٤) طالباً هو (٠,٦٨)، حدود الثقة عند ٩٥٪ هي (٠,٥٢) و (٠,٧٩). هذا يعني بأنّ الاحتمالية هي ٩٥٪ أن تكون قيمة الارتباط في مدى بين (٠,٥٢) و (٠,٧٩)، ويصف الباحث هذه النتيجة كالتالي: الارتباط بين الطول والوزن كان (٠,٦٨) ، وحدود الثقة عند ٩٥٪ هي (٠,٥٢) إلى (٠,٧٩)، ويرى لان(٢٠٠٠) بأنّ الحدود العليا والحدود الدنيا يمكن أن تفسر بشكل مستقل، الحدود الدنيا توضح مقدار صغر التأثير الموجود في المجتمع، فإذا كانت حدود الثقة لا تحتوي على الصفر (كما هو في المثال السابق)، فإنّ الارتباط دال إحصائياً، وهي طريقة أخرى للقول بأنّ الارتباط في المجتمع من غير المحتمل أن يكون صفراً، أو بمعنى آخر أنّ الارتباط الذي وجد في العينة ليس من نوع الارتباط المتوقع أن يراه الباحث إذا لم يكن هناك ارتباط في المجتمع (حسب الفرضية الصفرية)، الشكل (٦) يوضح العلاقة بين معاملات الارتباط وقيم (r) (Lane,2000).



شكل ٦ : حدود الثقة لعدّة ارتباطات عند مستويات مختلفة من **p**

يتضح من الشكل السابق أنّ حدود الثقة عند مستويي (r) (1.0) و(0.7) لا تظهر الدلالة الإحصائية حيث أنّ حدود الثقة تتضمن القيمة (صفر)، بينما أظهرت حدود الثقة الدلالة الإحصائية عند مستويي (r) (0.05) و (0.007) ، حيث أنّ حدود الثقة لم تتضمن القيمة (صفر).

وكما سبق أن ذكر لان Lane(2000) بأنّ قيمة الحد الأدنى من حدود الثقة توضح مقدار الحد الأدنى للتأثير أو الارتباط الممكن في المجتمع، ولذا فإنّ قيمة الارتباط في المثال الذي وضحه الشكل (٦) وإن كان دالاً إحصائياً عند مستوى (٠,٠٥) إلاّ أنّ ذلك الارتباط ضئيل جداً ولا يكاد يذكر، وقد وضح لان تأثر حدود الثقة بحجم العينة حسب الشكل التالي (Lane,2000):



شكل ٧ : العلاقة بين حدود الثقة وحجم العينة

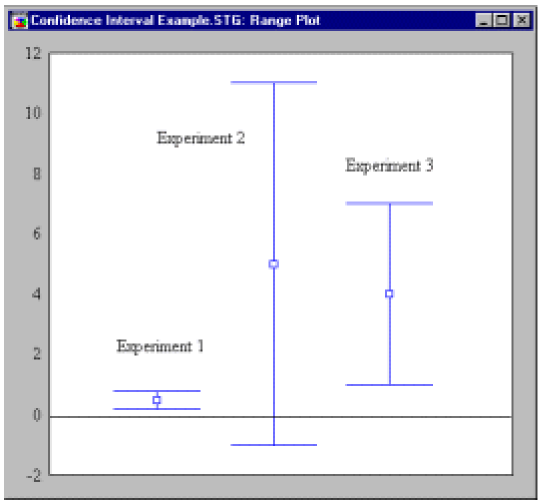
يتضح من الشكل (٧) أثر حجم العينة على مدى حدود الثقة، كما يتضح أيضاً أنّ الدلالة الإحصائية متأثرة بحجم العينة حيث لم تظهر تلك الدلالة الإحصائية من خلال حدود الثقة إلاّ عندما كان حجم العينة n يساوي (٦٤) وأكثر (Lane,2000). إنّ مثل المثالين السابقين اللذين يظهرا أنّ تقدير حدود الثقة يعطي نفس المعلومات التي تعطيها اختبارات الدلالة الإحصائية وأكثر، جعل ناب Knapp(1998) يصرّح بأنّه يفضّل حدود الثقة على اختبارات الدلالة الإحصائية.

لقد أشار لان Lane(2000) إلى فائدة بحثية أخرى لحدود الثقة تنفرد بها عن اختبارات الدلالة الإحصائية حيث رأى بأنّ الباحث إذا أراد إثبات أنّ معالجة ما ليس لها تأثير فلن يتمكن من ذلك باستخدام اختبارات الدلالة الإحصائية طالما أنّ عدم رفض الفرضية الصفرية لا يعني قبولها؟ فعلى سبيل المثال، وبافتراض أنّ أحد الباحثين يريد أن يثبت عدم تأثير عقار ما، وأنّ متوسطي المجموعة التجريبية والمجموعة الضابطة هما على التوالي m_1 و m_2 ، وأنّ تقدير حدود الثقة للمعلم ($m_1 - m_2$) عند مستوى الثقة ٩٩٪ كان:

$$-0.5 \le m_1 - m_2 \le 1.0$$

تشير النتيجة السابقة إلى أنّ الباحث سيكون متأكداً بدرجة ٩٩٪ أنّ فرق المتوسط الحقيقي لتأثير معالجة العقار هو في مكان ما بين (–٠,٥ و ١,٠)، فإذا كانت القيمة تساوي (–٠,٥) فإنّ العقار ضار بدرجة ما، أمّا إذا كانت القيمة تساوي (١) ، فإنّ العقار مفيد إلى حد ما، ولكن ما مقدار فائدة هذا العقار إذا كان متوسط الفرق في أحسن أحواله يساوي (١)؟ إجابة هذا السؤال تحتاج إلى سلّم المقياس، فإذا كان المقياس متكون من (٥٠) نقطة، والانحراف المعياري هو (١٠)، فهذا يعني أنّ أعلى مقدار للفرق والذي يساوي (١) يشكل ٠,١ من الانحرافات المعيارية، ممّا يعني أنّ التأثير صغير جداً، وهذا يعني أنّ نتيجة حدود الثقة توضح أنّ الفرق الأعلى الذي يمكن أن يتوقعه الباحث فرق صغير جداً، بحيث يستطيع الباحث أن يستنتج أنّ العقار غير فعال (Lane,2000).

إنّ التساؤلات الأساسية في الأبحاث هي: ما هو أفضل تخمين لحجم التأثير في المجتمع؟ وما مدى الدقة في تحديد حجم التأثير في المجتمع؟ اختبارات الدلالة الإحصائية تخفق في الإجابة عن هذه التساؤلات بشكل مباشر، بينما حدود الثقة تتضمن معلومات حول الدقة التجريبية التي لا تتوفر في نتائج اختبارات الدلالة الإحصائية، ولذا فإنّ حساب حدود الثقة أكثر فائدة من قيمة (r) لاختبار الفرضيات الصفرية، والمثال التالي يوضح مقارنة نتائج ثلاث دراسات:



شكل ٨ : مقارنة حدود الثقة لثلاث معالجات

الشكل السابق يزود الباحث بالعديد من الاستنتاجات منها:

- أنّ المعالجتين (١) و (٣) دالتان إحصائياً، بينما المعالجة (٢) غير دالة إحصائياً لأنّ تقدير حدود الثقة لها يتضمن الصفر.
- مع أنّ المعالجة (١) دالة إحصائياً إلاّ أنّ الفروق فيها (تأثير المعالجة) ضئيلة جداً، حيث أنّ الحد الأعلى للمعالجة هي قيمة صغيرة مقارنة بالحد الأدنى لقيمة الفروق في المعالجة (٣)، مما يعني أنّ المعالجة في التجربة (١) دالة إحصائياً لكنّ تأثيرها تافه وضئيل وغير مهم.
- المعالجة (٣) أكثر تأثيراً حيث يبدو أنّ الحد الأدنى لحدود الثقة مقبولة إلى حد ما مقارنة بالحد الأعلى لتأثير المعالجة (١)
- اعتماد الباحث على قيم (r) الناتجة من اختبارات الدلالة الإحصائية لن تمنحه الفرصة لتفسير نتائج التجربتين (١) و (٣) كما تمنحه حدود الثقة، فباستعمال قيم (r) المحسوبة سيستنتج الباحث أنّ التجربتين (١) و (٣) ذات دلالة إحصائية، ولكن لن يلاحظ الباحث الفرق بين الدلالتين العمليتين للتجربتين كما اتضح سابقاً من تفسير نتيجة التجربة (١) (StatSoft,2000).

ثالثاً – حجم التأثير Effect Size

مفهوم حجم التأثير

يرى ولكنسون (Wilkinson(1992 أنّ من النادر أن يبحث الباحث عن حجم التأثير، وأنّه غالباً ما يستنتج وبشكل خاطئ أنّ الدراسة لا تكشف عن أي تأثير حقيقي إذا لم يتم الوصول إلى الدلالة الإحصائية، وقد وصف هذا الوضع شيفر Shaver بقوله أنّ حجم التأثير قضية مهمة ولكن نماذج البحث التربوي لا تجعل الباحثين التربويين يفكرون في المقدار الذي يجب على حجم التأثير الوصول إليه لتعتبر النتائج مهمة، وبدلاً من ذلك تكيف الباحثون حالياً على تحديد مستوى الدلالة (a)، وإعطاء نتيجة اختبار الدلالة الإحصائية الأهمية القصوى، بينما يمنحون معالجة الدلالة التربوية أو العملية وضعاً ثانوياً. ولقد صرحت الجمعية النفسية الأمريكية APA في دليلها لعام ١٩٩٤م بأنّه "ليس أحد النوعين لقيم الاحتمالية [في اختبارات الدلالة الإحصائية] يعكسان أهمية أو مقدار التأثير بسبب أنّ كليهما يعتمدان على حجم العينة... ولذا فإننا نشجعك على تقديم معلومات عن حجم التأثير"، كما عبّر ماكلين وإرنست (McLean and Ernest(1998 عن توصلهما إلى تشجيع أدبيات البحث التربوي لمؤشرات حجم التأثير بقولهما: "في مراجعة الأدبيات، لم يجد المؤلفان مقالاً يجادل ضد قيمة تضمين بعض النماذج لحجم التأثير أو تقدير الدلالة العملية في تقارير الأبحاث"، بينما عبّر شيفر Shaver عن أهمية مقاييس حجوم التأثير بقوله: "أعتقد بأنّ الدراسات يجب أن تنشر بدون اختبارات الدلالة الإحصائية، ولكن ليس بدون أحجام التأثير" (McLean and Ernest,1998,p.18).

حجم التأثير هو الذي يقيس قوة العلاقة (التلازم) بين المتغيّرات الموجودة في الدراسة، وقد أوضح هايز أنّ مستوى الدلالة الإحصائية بمفردها لا تخبرنا بشيء عن قوة التلازم بين المتغيّرات، وكما قال ثومبسون بأنّ حجم التأثير هو الذي يوجه التفسير نحو ما هو مهم حقيقة في البحث ويجلب الانتباه لقضية جدارة النتائج (Wilkinson,1992).

ويرى كيلو (Kellow(1998 أنّ أول ما يميّز حجوم التأثير هو إمكانية تفسيرها بشكل مستقل عن حجم العينة، وهي مفيدة في تحديد المقدار الفعلي للفروق بين متوسطات المجموعات أو درجة التلازم بين المجموعات، ولذا فإنها تزود الباحثين بمعلومات تتجاوز حدود اختبارات الدلالة الإحصائية، وتمنحهم فهماً واضحاً للبيانات التي بين أيديهم. وقد أوضح ماكلين (McClain(1995 فائدة حجم التأثير في منع سوء تفسير الاختبارات الإحصائية بقوله "حجوم التأثير تمنح صورة أوضح للعلاقات الموجودة في البيانات، هذه المؤشرات ليست ضبابية ولا محجوبة بحجم العينة مثل الإحصاءات الناتجة من اختبارات الدلالة الإحصائية...

نحن نشعر بأنّ فهم واستخدام مؤشرات حجم التأثير يمنع الاختبارات الإحصائية من سوء التفسير كمؤشرات للأهمية ... بالتأكيد البحث النفسي يمكن أن يفسر بطريقة أفضل باستخدام حجوم التأثير" (p.299).

ولقد أوجز هوستون Huston(1993) فوائد مقاييس حجم التأثير على النحو التالي:

١ - حجم التأثير يشير إلى درجة وجود الظاهرة في المجتمع بمقياس متصل، بحيث يعني الصفر عدم وجود الظاهرة.

٢ - تزود الباحثين بمؤشرات للدلالة العملية بخلاف اختبارات الدلالة الإحصائية.

٣ - يمكن استخدامها في المقارنة الكمية بين نتائج دراستين أو أكثر كما هو مستخدم في التحليل البعدي Meta Analysis.

٤ - يمكن استخدامها في تحليل القوة الإحصائية لتحديد كم عدد العناصر المطلوبة في دراسة معينة.

تعريف حجم التأثير

يرى نيكس وبارنيت Nix and Barnette(1998) أنّ حجم التأثير (وبدون أي تضمين للسببية) هو الدرجة التي توجد فيها الظاهرة في المجتمع، أو الدرجة التي تكون فيها الفرضية الصفرية خاطئة، بينما يرى كيلو Kellow(1998)، وسنايدر ولاوسون Snyder and Lawson(1992)، وهوستون Huston(1993) بأنّ حجم التأثير هو ببساطة أي مقياس يخبر عن مدى تفسير المتغيّر التابع أو توقعه بواسطة المتغيّر المستقل.

يكشف التعريفان السابقان عن وجود صنفين عريضين لحجوم التأثير هما: مقاييس حجم التأثير Effect size التي تتطلب الفروق بين المتوسطات ، ومقاييس قوة التلازم Strength of association التي تستلزم نسبة التباين (مقدار التباين في المتغيرات التابعة المتلازمة مع التباين في المتغيّرات المستقلة (Huston,1993; Kirk,1996; Snyder and Lawson,1992).

أنواع حجم التأثير

لاحظ بعض الباحثين مثل سنايدر وثومبسون Snyder and Thompson(1992) بأنّ هناك عدة مصطلحات تستخدم للإشارة إلى مقاييس حجم التأثير منها:

(أ) مقاييس مقدار التأثير. (ب) مقاييس مقدار التأثير التجريبي.

(ج) مقاييس التباين المفسّر. (د) مقاييس قوة التلازم Association

(هـ) مقاييس قوة العلاقة.

لكنّ كيرك Kirk(1996) رأى بأنّ تلك المقاييس لحجوم التأثير يمكن أن تصنّف في صنفين رئيسين هما :

(أ) مقاييس الفروق وهي المشهورة بحجم التأثير.
(ب) مقاييس تفسير التباين.

أولاً- مقاييس حجم التأثير Effect Size

يرى كيلو Kellow(1998) بأنّ إحدى طرق التحقق من الأهمية العملية وقياس حجم التأثير هي فحص متوسط الفروق بين المجموعات لأنّ هذا الإجراء يجيب على التساؤل الذي من قبيل (ما مقدار الفرق الموجود بين المجموعات المبني على وحدة قياس المتغيّر التابع؟)، أمّا الطريقة الثانية فهي حساب المتوسط المعياري للفروق، حيث تعبّر هذه الطريقة عن تأثيرات المعالجة بوحدات الانحراف المعياري. يضرب كيلو مثالاً على الطريقتين على النحو التالي:

افترض أنّ هناك معالجة نفسية صممت لرفع مفهوم الذات لدى الشباب في سجن أحداث، تم تعيين الطلاب عشوائياً لاستقبال المعالجة التجريبية والمعالجة المقارنة، وبعد ستة أشهر من المعالجة كان متوسط المجموعة التجريبية في درجات مفهوم الذات (٤٢) نقطة مقارنة بالمجموعة المقارنة التي متوسطها (٤٠) نقطة، فإذا كان مقياس مفهوم الذات المعياري (من ١ إلى ١٠٠)، فإنّ هاتين الدرجتين اللتين هما الفرق العائد من المعالجة التجريبية ليستا مؤثرتين، علماً بأنّ هذا الفرق ربما يكون دالاً إحصائياً لو أنّ الدراسة أجريت على عدد كبير وكافٍ من العناصر، بغض النظر عن أهمية النتائج النهائية (Kellow,1998).

الطريقة الثانية لقياس حجم التأثير ومن ثمّ تفسير الدلالة العملية هي حساب المتوسط المعياري للفروق، وهي طريقة تعبّر عن تأثيرات المعالجة بوحدات الانحراف المعياري، ولذا فإنّ الفروق يمكن أن تفسّر باستخدام النسب المئوية للتحسّن باستخدام المنحنى الطبيعي

المعياري، فعلى سبيل المثال، إذا أراد باحث تقويم برنامج جديد للقراءة باستخدام مقياس (١ - ١٠٠)، وبعد ستة أشهر من التدريس، كان متوسط مجموعة المعالجة التجريبية (٧٠) نقطة، بينما كانت المجموعة المقارنة للطلاب الذين استقبلوا الطريقة التقليدية في التدريس هي (٦٠) نقطة، وبافتراض أنّ الانحراف المعياري لكل مجموعة هو (١٠) نقاط، فإنّ الفروق المعيارية بين المجموعتين هي (١٠) نقاط مقسومة على الانحراف المعياري أي ($\frac{10}{10}$)، هذه النتيجة في الفرق تساوي انحراف معياري واحد. لقد سمى كوهين (١٩٨٨) هذه الإحصاءة (d) ، وعند الرجوع إلى جدول التوزيع الطبيعي المعياري نجد أنّ متوسط درجات المجموعة التجريبية (والتي تساوي في هذا المثال انحراف معياري واحد فوق الطريقة التدريسية التقليدية) هو ٣٤٪ أعلى من مجموعة الطريقة التقليدية في التجربة. التفسير لهذه الفروق مباشر جداً وهو أنّ ٣٤٪ من الطلاب الذين تعرضوا للطريقة الجديدة حصلوا على درجات أعلى من الذين تعرضوا للطريقة التقليدية، ويلاحظ أنّ درجات الفروق المعيارية تخبر الباحث بشيء مختلف عمّا تخبر به مقدار الدرجات الخام (الطريقة الأولى)، ومنها يمكن التعرّف على عدد العناصر الإضافيين الذين استفادوا من البرنامج (Kellow,1998).

ولاستخدام مؤشرات الدلالة العملية (حجم التأثير) المرافقة لاختبارات الدلالة الإحصائية الشائعة يمكن الرجوع إلى كتابي كوهين (1988)Cohen والقرشي (١٤٢١) حيث عرضا أهم المؤشرات ومنها:

(أ) مؤشر حجم التأثير (d) لاختبار (ت) للفروق بين المتوسطات:

قدّم كوهين (1988)Cohen مؤشر (d) لحجم التأثير لنتائج اختبار (ت) حسب المعادلة التالية:

$$d = \frac{|m_1 - m_2|}{s}$$

حيث : (d) هو مؤشر حجم التأثير.

($m_1 - m_2$) هو الفرق بين متوسطي العينتين.

(s) هو الانحراف المعياري لإحدى العينتين (بافتراض تساويهما).

أمّا إذا كان الانحراف المعياري للعينتين غير متساويين، فإنّ قيمة الانحراف المعياري المستخدم في المعادلة السابقة يحسب كالآتي (Cohen,1988):

$$s = \sqrt{\frac{(s_1)^2 + (s_2)^2}{2}}$$

حيث : (s_1) و (s_2) هما الانحراف المعياري للعينتين (١) و (٢).

ويمكن الحصول على المؤشر (d) من نتيجة اختبار (ت) على النحو التالي:

$$d = 2t\sqrt{dfs}$$

ومن خلال قيمة (d) يمكن حساب معامل ارتباط بيرسون (r) على النحو التالي:

$$r = \frac{d}{\sqrt{d^2 + 4s}}$$

حيث يعبّر مربع معامل الارتباط (r^2) عن نسبة التباين المشترك الذي يمكن أن يرجع إلى أحد المتغيرين إذا كان التصميم ارتباطياً، وعن نسبة التباين الذي يرجع للمتغيّر المستقل إذا كان التصميم تجريبياً (القرشي، ١٤٢١، ص٣٥٣).

(ب) مؤشر حجم التأثير (d) لاختبار (ت) للفروق بين المتوسطات:

يستخدم معامل الارتباط الثنائي الأصيل (r_{pbis}) كمؤشر لحجم التأثير في اختبار (ت) حسب المعادلة التالية:

$$r_{pbis} = \sqrt{t^2/(t^2 + df)}$$

ويعبّر مربع معامل الارتباط الثنائي الأصيل (r^2_{pbis}) عن نسبة التباين في درجات المتغيّر التابع التي يمكن إرجاعها إلى المتغير المستقل إذا كان التصميم الذي استخدم فيه اختبار (ت) تجريبياً، أمّا إذا كان التصميم ارتباطياً فإنّ مربع معامل الارتباط الثنائي الأصيل (r^2_{pbis}) يعبّر عن نسبة التباين المشترك التي يمكن أن ترجع إلى أي من المتغيرين (القرشي، ١٤٢١، ص٣٥٢).

(ج) مؤشر حجم التأثير (f) لاختبار تحليل التباين (ف) :

$$f = \sqrt{\frac{h^2}{h^2 - 1}}$$

حيث : (h^2) هي مربع ايتا ، وتحسب كالآتي:

$$eta^2 = h^2 = \frac{SS_{between}}{SS_{total}}$$

حيث: $SS_{between}$ هو مجموع المربعات بين المجموعات.

SS_{total} هو مجموعات المربعات الكلي (Cohen,1988).

(د) مؤشر حجم التأثير (W) لاختبار مربع كاي (c^2) :

$$W = \sqrt{\frac{c^2}{n}}$$

حيث: (n) هي حجم العينة (Cohen,1988).

(ه) مؤشر حجم التأثير الخاص بمعامل ارتباط بيرسون (ر) :

مؤشر حجم التأثير الخاص بمعامل ارتباط بيرسون (ر) هو (ر) نفسه، ويستخدم للارتباط بين متغيرين متصلين، كما يعبّر مربع معامل ارتباط بيرسون (ر٢) (r2) عن نسبة التباين المشترك الذي يمكن أن يرجع إلى أحد المتغيرين (القرشي، ١٤٢١،ص٣٥١).

كما أنّ كوهين قد وضع جداول خاصة لاستخراج جميع المؤشرات السابقة ومؤشرات أخرى أيضاً (Cohen,1988).

ثانياً : تقدير نسبة التباين المفسّر

الطريقة الثالثة للتحقق من الدلالة العملية للنتائج الدالة إحصائياً، هي تقدير نسبة التباين في المتغيّر (المتغيّرات) التابعة التي تفسّر بواسطة المتغيّر (المتغيّرات) المستقلة، ولذا فإنّها عبارة عن نسبة التباين المفسّر إلى التباين الكلي(Snyder and Lawson,1992). وهناك مؤشرات كثيرة لتقدير التناسب في التباين المفسّر منها نسبة الارتباط المعروفة مربع إيتا eta2 (h^2) ، ومربع أوميغا Omega2 (w^2) ، ومربع إيبسلون (e^2) وغيرها من المؤشرات، وبشكل عام فإنّ جميع طرق الاختبارات البارامترية (المعلمية) الشائعة يمكن استخدام تقديرات نسبة التباين المفسر بسبب أنّها ارتباطية Correlational، وبسبب ذلك فإنّ مقياس قوة التلازم (الترابط) مثل (R^2) الناتجة من اختبار الانحدار يمكن أن تستخلص من الطرق المعلمية الأخرى، فعلى سبيل المثال القيمة المناظرة لـ (R^2) في ANOVA هي eta2 (h^2)،

والتي هي ببساطة عبارة عن: $eta^2 = h^2 = \frac{SS_{between}}{SS_{total}}$ (Kellow,1998). ويرى الباحث بأنّ مقال كيرك Kirk(1996) يعد مرجعاً جيداً للمهتمين بمؤشرات الدلالة العملية لأنّه يتضمن مؤشرات عديدة.

مشكلات مقاييس حجم التأثير

مع أنّ مقاييس حجم التأثير مهمة في تفسير نتائج التحاليل الإحصائية إلاّ أنّها مثل أي تقنية إحصائية لا يخلو استخدامها من المشكلات. ويرى الباحث بأنّ الاعتقاد المفرط في فائدة هذه المقاييس دون التنبه إلى حدودها قد يجعل الباحثين التربويين يقعون في مشكلات جديدة نتيجة خضوعهم لأسر مقاييس حجم التأثير تماماً مثلما حدث عند خضوعهم لأسر اختبارات الدلالة الإحصائية، ولذا عرض الباحث وباختصار موجز أهم مشكلات مقاييس حجم التأثير وهي:

١ - أغلب مؤشرات مقدار التأثير مثل الفرق المعياري، وتناسب التباين المفسّر، شديدة التأثر بعدم تجانس العينة Heterogeneity ، ولذا فإنّها لا تقيس المقدار الحقيقي لتأثير المعالجة بنفس وحدة قياس المتغيّر التابع، ولقد عرض كيلو(١٩٩٨) مثالاً يوضح هذه المشكلة حسب الجدول التالي(Kellow,1889):

جدول ١١ : تأثر مقاييس نسبة التباين المفسّر بمدى تجانس وعدم تجانس الدرجات

المعالجة (١)		المعالجة (٢)	
الضابطة ($n=4$)	المعالجة ($n=4$)	الضابطة ($n=4$)	المعالجة ($n=4$)
٤	٧	٢	٥
٥	٨	٤	٧
٥	٨	٦	٩
٦	٩	٨	١١

ملخص إحصاءات المعالجتين السابقتين هو على النحو التالي:

المعالجة (١)	المعالجة (٢)
متوسط المجموعة الضابطة = ٥,٠	متوسط المجموعة الضابطة = ٥,٠
متوسط مجموعة المعالجة = ٨,٠	متوسط مجموعة المعالجة = ٨,٠
الانحراف المعياري للضابطة = ٠,٨٢	الانحراف المعياري للضابطة = ٢,٦
الانحراف المعياري للمعالجة = ٢,٦	الانحراف المعياري للمعالجة = ٠,٨٢
Eta^2 = ٠,٨٢	Eta^2 = ٠,٣١
d (مؤشر كوهين) = ٣,٦	d (مؤشر كوهين) = ١,٢
$r_{calculated}$ = ٠,٠٠٢	$r_{calculated}$ = ٠,١٥

من خلال الإحصاءات السابقة، يلاحظ ما يلي:

(أ) المعالجة الأولى قيمة ($r_{calculated}=0.002$) منخفضة جداً، فهي دالة إحصائياً، مقارنة بالمعالجة الثانية التي فيها ($r_{calculated}=0.15$) أي أنّها غير دالة إحصائياً.

(ب) إذا كان الباحث مدركاً لمفهوم الدلالة الإحصائية ومشكلاتها فسيدرك بأنّ قيمة الدلالة الإحصائية لا تعني الأهمية أو الدلالة العملية، وسيدرك أنّ عليه أن ينظر في مؤشرات الدلالة العملية (مقدار التأثير) لأنّها هي المعنية بذلك.

(ج) عند النظر في الجدول السابق يلاحظ أنّ مؤشرات الدلالة العملية سواء مؤشر الفرق المعياري (d) ، أو تناسب التباين المفسّر (Eta^2) تشير أيضاً إلى أنّ التأثير في المعالجة الأولى أكبر من التأثير في المعالجة الثانية، ولذا قد يشعر الباحث بالاطمئنان بأن يقرّر بأنّ المعالجة الأولى أكثر تأثيراً وفاعلية من المعالجة الأولى.

ولكن يجب أن يلاحظ من خلال البيانات والإحصائيات السابقة ملاحظة هامة، وهي أنّ مقدار الفرق الخام بين متوسطي المجموعة الضابطة والمجموعة التجريبية في المعالجة الثانية تساوي ٣,٠ ، وهي نفس القيمة تماماً في المعالجة الأولى. هذا يعني بأنّ مؤشرات الدلالة العملية المستعملة في الحسابات السابقة لا تقيس المقدار الحقيقي لتأثير المعالجة، وأنّ تلك المؤشرات متأثرة بشكل كبير بمقدار عدم التجانس بين البيانات في المعالجة الأولى، وهذا يعني بأنّه عندما تكون درجات المتغيّر التابع متفاوتة (متقلبة) بشكل كبير فإنّ التأثيرات المتماثلة ربما تحجب بواسطة تقديرات نسبة التباين المفسر (Kellow,1998). ربما يستخلص الباحث ممّا سبق ما أقرّه ماكسويل وآخرون بأنّ الباحثين يجب أن يكونوا حذرين حتى لا يصبحوا مرّة أخرى أسرى لمقياس التلازم Association كما حدث في مستويات (r) (Kellow,1998).

٢ - مقاييس التلازم تعتمد على المفهوم الرياضي الواسع الذي يعمل في نماذج التحليل الخطي، وهذا يعني أنّ هناك تحيّزاً إيجابياً وإفراطاً في النتائج المقدّرة Overestimates ، وقد لاحظ جراديو بأنّ مقدار التحيّز في مقاييس قوة التلازم تعتمد على عدّة عوامل منها: ثبات بيانات القياس Reliability ،

والتساؤلات البحثية المثارة، وحجم العينة (n) ، وعدد المتغيّرات المستقلة، وعدم تجانس عينة الدراسة Heterogeneity ، ونوع التصميم المستخدم في فحص الدراسة، وهذا جعل العلماء يطورون ما أسموه بالمقاييس غير المتحيّزة Unbiased لتقديرات قوة التلازم، وذلك في محاولة للتصحيح الإحصائي للتحيّز في المؤشرات المستعملة (Snyder and Lawson,1992). ومع ذلك حتى معادلات التصحيح تنزع إلى أن تكون أكبر عندما تكون حجوم العينة صغيرة، أو إذا كان حجم التأثير صغيراً أيضاً (Palomares,1990).

٣ - نظراً لأنّ مقاييس مقدار التأثير تتفاوت بسبب عوامل إحصائية، وعوامل طرق بحثية، فإنّ قضية (التأثيرات الكبيرة) مقابل (التأثيرات الصغيرة) تصبح محيّرة ومربكة، وتجادل الباحثون في قضية (ما مقدار) أو (ما حجم) مقياس التأثير الذي يجب أن يتحقّق حتى نعتبر النتائج ذات قيمة وفائدة؟ وعلى الرغم من أنّ كوهين (Cohen(1988 زود الباحثين ببعض التعليمات حول تقويم التأثيرات (كبيرة، وسط ، صغيرة)، فإنّ الحكم المتعلق بالدلالة العملية لمقدار التأثير يعتمد على نظام القيم لدى الباحث، وحسب التساؤلات الموضوعة، والاعتبارات الاجتماعية، وتصميم الدراسة، أي أنّه يعتمد على سياق الدراسة. وهذا يعني بأنّ قياسات حجم التأثير محدّدة حسب السياق وتتأثر بعوامل وأوضاع معيّنة في البحث، ولذا فإنّها مثل اختبارات الدلالة الإحصائية لها حدود معيّنة وتحيّزات، ولذا فإنّ الباحثين الذين يعرفون فوائدها وحدودها ووسائل التفسير المساعدة الأخرى، أكثر احتمالية لعمل استخدامات مفيدة لمقياس حجم التأثير في تفسير النتائج الجوهرية (Snyder and Lawson,1992).

٤ - حجوم التأثير يمكن إساءة تفسيرها تماماً مثل إساءة تفسير نتائج اختبارات الدلالة الإحصائية، فالتقديرات العشوائية التي وضعها كوهين لحجوم التأثير (منخفض، متوسط، وعالي) (مثل ٠,٢، ٠,٥، و ٠,٨ على التوالي) لفروق المتوسط المعيارية لا تعكس القيمة الحقيقية للدراسة. وقد لاحظ بعض الباحثين مثل قلاس وماك قرو وسميث بأنّ حجم التأثير الذي قيمته (d =1.0) ربما يعكس نتيجة تافهة ، فربّما يكون المتغيّر التابع ذا فائدة أو قيمة قاصرة جداً، أو ربما تكون النتيجة مكلفة جداً للإنتاج، أو أن يكون الثبات مشكوكاً فيه، وباختصار فإنّ استبدال مستوى ٠,٠٥ للدلالة الإحصائية المكرّسة حالياً، بتكريس تقليد لحجم التأثير لا يعتبر تقدماً (Shaver,1992).

٥ - مؤشرات الدلالة العملية لا تغني عن الدلالة الإحصائية، وسبب ذلك أنّ مؤشرات الدلالة العملية لا تزود الباحث بأي إثبات أو برهان على احتمالية حدوث النتائج، فبافتراض أنّ تجربة ما تحتوي على عنصرين فقط، وأنّ مؤشرات الدلالة العملية كانت مفيدة وكبيرة، فإنّ هذه المؤشرات للدلالة العملية لا تساعد الباحثين في إثبات أنّ تلك النتائج المحصّلة لم تكن بسبب الصدفة، ولذا من الضروري تضمين الدلالة الإحصائية وتقديرات الدلالة العملية مترافقين (McLean and Ernest,1998).

٦ - هناك خلط أحياناً بين قيمة حجم التأثير التي يفترضها الباحث قبل الدراسة وذلك لتحديد حجم العيّنة المناسب، وبين حجم التأثير الذي يحصل عليه الباحث كمؤشر للدلالة العملية لنتائج العينة، ولقد حاول كوهين التفريق بين المؤشرين بوضع شرطة على (d) على نحو ($\bar{d}$) لحجم التأثير المحسوب، ويرى ناب Knapp(1998) أنّ بعض هذا الخلط في مناظرات اختبارات الدلالة الإحصائية يمكن تجنبه إذا كان لدى الباحثين مصطلحين مختلفين لهذين النوعين من حجوم التأثير.

٧ - مؤشرات حجم التأثير تواجه صعوبة في شرح حجم التأثير في حالة المتغيّرات المتعدّدة، وقد اعتبر كوهين بأنّ هذه المؤشرات لتلك المتغيّرات ليست سهلة وليست معروفة (Nix and Barnette,1998b).

٨ - ليست كل معادلات نسب التباين المفسّر تعطي نتائج متماثلة، ففي الحقيقة هناك عدّة قضايا يجب اعتبارها في تحديد أي المعادلات البديلة يجب استخدامها، ومن تلك العوامل التي تحدّد المعادلات المناسبة ما يلي :

(أ) التقديرات المتحيّزة مقابل غير المتحيّزة.

(ب) التعميم للعينات المستقبلية مقابل التعميم للمجتمع.

(ج) تقديرات النماذج عشوائية التأثير مقابل ثابة التأثير.

(د) تقديرات نسبة التباين المفسّر للمتغيّرات أحادية المتغيرات Univariate مقابل عديدة المتغيرات Multivariate.

يستخلص الباحث ممّا سبق ما قرّره موراي ودوسر بأنّه "لا يوجد بديل إحصائي لخبرة ومعرفة الباحث. هذه الخبرة والمعرفة يجب أن تكون الأساس الذي يقرّر على أساسه الباحث ما هو حجم الفروق المهمة، وما هي التأثيرات العملية في دراسة موضوع بحثي معيّن" (Snyder and Lawson,1992,p.22)، ولذا فإنّ قياسات حجم التأثير لا تخلي الباحثين من مسئوليتهم في تحديد معنى بياناتهم (Huston,1993).

رابعاً – تحليل الإعادة Replication

مفهوم الإعادة وأهميتها

إعادة النتائج، تعتبر من القضايا المهمة التي يجب أخذها بعين الاعتبار. فقد قرّر كارفر Carver بأنّ الإعادة هي حجر زاوية العلم، بينما ذكر ثومبسون Thompson بأنّ الإعادة هي الاختبار النهائي والأقصى لتعميم النتائج، وحيث أنّ الدلالة الإحصائية لا تعني بالضرورة بأنّ المعالجة فعّالة، بل تشير إلى ما إذا كانت أخطاء المعاينة هي تفسير بديل للنتائج حيث ربما كانت في الحقيقة هناك عدة أسباب لإنتاج تلك النتيجة مثل مهدّدات الصدق الداخلي، ولذا فإنّ تصميم البحث وإجراءاته هي التي تؤثر في الإعادة وهي التي لا تشرح عادة بشكل كافٍ في الدراسات (Wilkinson,1992).

النّاس في الحقيقة لا يريدون أن يعرفوا المعلومات أو النتائج التي تخص العينة والتي هي عبارة عن مجموعة من المشاهدات التي لن تؤخذ مرّة أخرى، فهم أكثر اهتماماً بالمجتمع الذي سحبت منه العينة، لأنّ الأشياء التي تكون صحيحة في المجتمع، سوف تكون صحيحة في أكثر وأغلب الأفراد، إنّ هذا هو الغرض الرئيس الثاني من الإحصاء وهو التعميم من العينة إلى المجتمع (Lane,2000).

وعلى الرغم من أنّه معروف وعلى نطاق واسع بأنّ اختبار الدلالة الإحصائية هو ناتج بشكل كبير لحجم العينة أكثر من أهمية النتيجة أو تعميمها، فإنّ الدلالة الإحصائية لا تزال وفي الغالب تخلط بالإعادة (Gillaspy,1996).ولقد ناقش كل من كوهين Cohen وثومبسون Thompson كيف أنّ اختبار الدلالة الإحصائية لا يقوم إعادة النتيجة، علماً بأنّ النتيجة التي تعاني من قصور في التعميم لها قيمة علمية محدودة بغض النظر عن الدلالة الإحصائية أو الدلالة العملية، ولذا يجب على الباحثين الحذر من وضع ثقة كبيرة في نتائجهم حتى يتأكدوا من إعادة النتائج (Gillaspy,1996)، فمعلومات الإعادة مهمة ومختلفة عن معلومات أهمية النتائج (McLean and Ernest,1998).

إنّ العلم يتقدّم بواسطة تراكم الأحداث والبراهين بأنّ نتائج معيّنة تحدث تحت نفس الظروف (Snyder and Thompson,1998). والعلم ببساطة هو عمل اكتشافات التأثيرات المتكرّرة (McLean and Ernest,1998). وبغض النظر عن الطريقة أو المؤشر المستعمل، فإنّ احتمالية الإعادة للدراسة معلومة مهمة للقارئ وفي الاهتمام الأول للعالم الجيّد (Nix and Barnette,1998b).

لقد ناقش سنايدر وثومبسون Snyder and Thompson(1998) الإعادة بشكل جيد، حيث قالا: إنّ كثيراً من الباحثين يمنحون الاختبارات الإحصائية أهمية مبالغ فيها بسبب اعتقادهم غير الصحيح بأنّ قيمة (r) تقوّم احتمالية أن تكون نتائج العينة تحدث في المجتمع (أو أنّ الفرضية الصفرية خاطئة)، هذا الاستنتاج فيما يخص المجتمع مهم جداً، لكن الدلالة الإحصائية لا تختبر قيم المجتمع، اختبار المجتمع مفيد جداً، إذا عرف الكثير حول المجتمع فإنّه سوف يعرف أكثر حول ما يمكن أن يجده الباحثون الآخرون في العينات المستقبلية المسحوبة من المجتمع، واختبارات الدلالة الإحصائية لا تحسب احتمالية نتائج المجتمع لنتائج عيّنة معطاة، فبدلاً ممّا سبق، فإنّ عدداً من الباحثين يركزون بأنّ اختبارات الدلالة الإحصائية تقوّم احتمالية قيم العيّنة وذلك بافتراض أنّ الفرضية الصفرية ممثلة بشكل تام للمجتمع، هذه قضية مختلفة... قيم المجتمع وقيم العينة، أي القيم مأخوذة كقيم معطاة! هناك اختلاف، وهذا يعني أنّ التفسيرين متعارضين ومتناقضين، وببساطة فاتجاه الاستدلال الإحصائي في اختبارات الدلالة الإحصائية هو من المجتمع للعينة، وليس من العينة للمجتمع (Snyder and Thompson,1998).

وكما وضح كوهين Cohen فإنّ اختبار الفرضية الصفرية المعروف لا يخبر الباحث عمّا يريد أن يعرف، ما يريد الباحث أن يعرفه هو أنّه بمعلومية بيانات العينة ما هي احتمالية أن تكون H_o صحيحة في المجتمع؟ ولكن ما يخبر به اختبار الفرضية الصفرية هو أنّه بمعلومية أنّ H_o صحيحة في المجتمع، ما هي احتمالية أن تكون بيانات العينة متطرفة جداً؟ هذان ليسا نفس الشيء (Snyder and Thompson,1998).

ومع أهمية دراسات الإعادة، إلاّ أنّ هناك عدّة عوامل تعمل ضدها، منها الازدراء العام للبحث غير الأصلي من قبل محرّري المجلات ولجان الأطروحات، كما أنّ نقص المعلومات في بعض الدراسات يخلق صعوبة في إعادتها، أمّا العامل الثالث فهو التحيّز الذي يحدث عندما يعيد الباحث نفس دراسته مرة أخرى بعد الانتهاء من إكمالها (Nix and Barnette,1998).

طرق تقويم الإعادة

هناك عدّة طرق لتقويم إعادة النتيجة، تتضمن هذه الطرق دراسة الإعادة الحقيقية (الخارجية) مع عينات جديدة، أو تحاليل الإعادة الداخلية التي تحاكي عمليات الإعادة الحقيقية. ولقد أوصى كل من ثومبسون Thompson(1999c) وروبنسون وليفين (McLean and Ernest,1998) بأهمية تنفيذ الإعادة الخارجية، واعتبروا أنّ تحليل الإعادة الداخلية يشكل بديلاً مقبولاً، وأنّ دراسات الإعادة الخارجية هو المثالي في كل

الأحوال، ولكن نظراً لأنّ الباحثين ليست لديهم القدرة والاحتمال للإعادة الخارجية، فإنّ تحليل الإعادة الداخلية يساعد في تحديد فائدة النتائج.

تحاليل الإعادة الداخلية الأساسية ثلاثة هي :

- طريقة التحقق المتبادل للصدق (Cross-Validation).
- الطريقة الركمية المنتهية باختيار متعدّد (The bootstrap) .
- الطريقة السبرية المنتهية باختيار مفرد (The jackknife)

(Thompson and Snyder,1997a; Daniel,1997; Nix and Barnette,1998).

تتطلب الإجراءات الثلاثة السابقة تنفيذ التحليل على مجموعات جزئية مختلفة لمجموعة البيانات الأصلية، ومن ثمّ مقارنة تلك النتائج المختلفة (Gillaspy,1996; Daniel,1998a).

أ) طريقة التحقق المتبادل للصدق (Cross-validation):

تعتبر هذه الطريقة الأبسط لتقويم إعادة النتائج، وتطبّق هذه العملية على ثلاث خطوات هي:

1) يجزئ الباحث وبشكل عشوائي العينة إلى مجموعتين.
2) ينفذ الباحث الاختبار الإحصائي المستعمل في الدراسة على المجموعتين الجزئيتين بشكل مستقل.
3) يقارن الباحث امبريقياً النتائج لمعرفة إعادة النتائج للمجموعتين الجزئيتين، ممّا يزيد الثقة في إعادة نتائج الدراسة، ويمكن ذلك من خلال حساب معامل الثبات المحصّل حيث يجب أن يقترب من الواحد الصحيح.

ومع ذلك فإنّ تفسير النتائج يجب أن يعتمد دائماً على العينة الكلية، وليس على أجزاء المجموعات الجزئية. والسبب أنّ العينة الكاملة – نظريا - يجب أن تعطي النتائج الأكثر قابلية للتعميم، بينما تجزئة العينة تجرى فقط لتقويم إعادة النتائج (Palomares,1990). إيجابية طريقة التحقق المتبادل للصدق (Cross-Validation) أنّها تتطلب عينة بيانات واحدة، وسلبيتها أنّ تقسيم البيانات الأصلية يقلّل حجم العيّنة، وهذا يخلق مشكلة إذا كانت العيّنة أصلاً صغيرة(Gillaspy,1996; Nix and Barnette,1998).

ب) الطريقة الركمية المنتهية باختيار متعدّد (Bootstrap)

تعتبر هذه الطريقة واحدة من أقوى طرق تقويم إعادة النتائج حيث تعتمد الطريقة على نسخ البيانات الأصلية فوق بعضها عدد كبير من المرّات وذلك لخلق ملف ضخم من البيانات (Mega File)، ومن هذه المجموعة الضخمة، تختار مئات أوآلاف من العينات العشوائية بحيث يكون حجم كل عينة هو نفس العدد (n) للعينة الأصلية، ويجري الاختبار المطلوب والمحدّد في نفس الدراسة، ثم تحسب نتائج تلك الاختبارات لكل عينة بشكل مستقل، ويؤخذ المعدّل (Palomares,1990). وقد طوّر ثومبسون عام ١٩٩٢م برنامج حاسب آلي لتنفيذ الخطوات السابقة بشكل آلي، سمّاه DISCSTRA (Gillaspy,1996).

ج) الطريقة السبرية المنتهية باختيار مفرد (Jackknife)

طورت هذه الطريقة بواسطة توكي Tukey عام ١٩٥٨م (Gillaspy,1996). تختلف هذه الطريقة السبرية (Jackknife) عن الطريقة الركمية (Bootstrap) بأنّ المجموعات الجزئية المختلفة أو مجموعات العناصر تسحب بشكل متكرّر من خارج المجموعة الأصلية للبيانات، والطريقة هي على النحو التالي:

١) يجري الباحث الاختبار المستعمل في الدراسة على البيانات الأصلية.
٢) يقسم الباحث العينة الأصلية (N) إلى عدد (k) من المجموعات الجزئية متساوية الحجم (n)، بحيث كل مجموعة جزئية يمكن أن تكون صغيرة إلى حد (عنصر واحد فقط) أو كبيرة إلى أكبر حد مناسب.
٣) يمسح الباحث وبشكل متكرّر كل مجموعة جزئية من العينة الأصلية.
٤) يحسب الباحث قيمة إحصاءة الاختبار المستخدم في الدراسة للمجموعة الجزئية والمجموعة التي سحبت منها المجموعة الجزئية، ويجب ملاحظة أنّه كلما كانت المجموعة الجزئية أصغر، نتجت إعادة أكثر، ممّا يجعل من الأسهل ملاحظة النتائج وتكون هناك ثقة أكبر فيها.
٥) يحسب الباحث معدّل الإحصاءة لإنتاج معامل الـ (Jackknife)، ثمّ يفسّر النتائج.

الفرق بين الطريقة السبرية (Jackknife) و الطريقة الركمية (Bootstrap) أنّ الأولى تعاد فيها المعاينات بدون إحلال، أمّا الثانية فإنّ المعاينات تعاد فيها مع الإحلال (Knapp,1998). كما أنّ مشكلة جميع الطرق الثلاث السابقة أنّها منحازة إلى حد ما حيث أنّها تميل لإنتاج تقديرات منحازة إيجابياً لإعادة النتيجة، وذلك لأنّ عينة واحدة هي المستعملة في كل طريقة، ومع ذلك فهي تمثّل أفضل بديل لإجراء اختبارات الإعادة لدراسة محدّدة (Daniel,1998b; Thompson and Snyder,1997a).

إجابة السؤال الثالث

الأساليب المقترحة لتفسير نتائج اختبارات الدلالة الإحصائية

يرى ماكلين وإرنست McLean and Ernest(1998) بأنة ليس إحدى التوصيات المقترحة بخصوص تحسين الممارسة في استعمال اختبارات الدلالة الإحصائية في البحث التربوي هي جديدة كلياً، ولكنّ التوصيات تزود بوجهة نظر واسعة عن الاعتراضات، وتزود بإرشاد عملي للباحثين لتوظيف اختبارات الدلالة الإحصائية في أعمالهم. ويرى هوستون Huston(1993) بأنّ الأزمات في اختبار الدلالة الإحصائية دفعت طرقاً إحصائيةً أخرى للأمام لمساعدة الباحثين في تحديد وتفسير أهمية نتائجهم. ويصف نيكس وبارنيت Nix and Barnette(1998) البحث عن البدائل بقوله "نحن لا نعتقد بوجود بديل متوفر الآن، ولكنّ الطرق والأساليب البديلة لديها القدرة والاستعداد إذا استخدمت بطريقة صحيحة لتحركنا خارج المستنقع الحالي كما وصفه ميهل قبل ٢٠ عاماً" (p.7). أمّا هوستون Huston(1993) فيرى بأنّ الاعتماد الأقل على اختبارات الدلالة الإحصائية وزيادة استخدام أساليب أخرى بشكل متلازم ربما يقودنا إلى تحسين نوعية الأبحاث الجديدة.

يتضح ممّا سبق عدم وجود حل بديل محدّد، إنّما هناك عدّة أساليب إحصائية يمكن أن تستعمل بطريقة أو بأخرى لتكميل فائدة اختبارات الدلالة الإحصائية، ولتحسين الممارسة الإحصائية في البحث التربوي، ويمكن صياغة تلك الاستراتيجيات على النحو التالي:

أولاً – إعادة صياغة الأسئلة البحثية والفرضيات

لقد اتضح من خلال تحديد مشكلات الدلالة الإحصائية في البحث التربوي في هذا الفصل وجود انتقاد من قبل العديد من الإحصائيين في أسلوب صياغة الفروض التي تبحث عن وجود أو عدم وجود فروق، وأنّ أولئك الإحصائيين يرون أنّ الفرضية الصفرية لعدم وجود الفروق مستحيلة أن تكون صحيحة في المجتمع لأنّه دائماً ما تكون هناك بعض الفروق في معالم المجتمع حتى لو كانت تلك الفروق ضئيلة، ولذا فإنّ الباحثين يعرفون مسبقاً وقبل جمع البيانات بأنّ الفرضية الصفرية عادة ما تكون خاطئة لأنّ أي دراسة ربما تظهر نتائج دالة إحصائياً إذا استخدم الباحث عناصر كافية في عينة البحث (Snyder and Thomson,1998).

يفترض أن يكون التساؤل الأساسي الذي يجب أن يجاب عليه عند تنفيذ البحث هو "ما مقدار مسؤولية المتغيّر المستقل عن المتغيّر التابع" أو "ما نسبة التباين في المتغيّر التابع الذي يفسّر بالتأثير المشاهد؟" (Keaster,1988; Huston,1993). وكما ذكر ثومبسون Thompson(1988) بأنّ التركيز على حجوم التأثير يجعل الباحث يركز على ما هو مهم حقاً مثل قضية فائدة النتائج وقيمتها.

ثانياً – فحص البيانات بشكل موسع

بغض النظر عن استعداد الباحثين لإعادة صياغة تساؤلات بحثهم أو لا فإنّ الأهم من ذلك هو استكشاف وتحرّي وفحص البيانات الذي يعتبر قضية مهمة يجب تشجيع الباحثين عليها أكثر من مجرّد الاعتماد على اختبار الدلالة الإحصائية. ويرى ثومبسون Thompson(1998c)، وماكلين وإرنست McLean and Ernest(1998)، وويلكنسون Wilkinson(1992) بأنّ من المناسب فحص البيانات بعدة طرق، حيث أنّ اختبار الدلالة الإحصائية تفيد فقط في فحص النتائج هل هي بسبب الصدفة، أمّا تقويم النتائج كيف هي مؤثرة فإنّها تحتاج إلى أساليب أخرى. ولذا يقترح ما يلي:

١ - أن يحدّد الباحث مسبقاً حجم التأثير المطلوب الكشف عنه فعلاً ثمّ يحدّد بناء عليه حجم العينة المطلوب الحصول عليها للوصول إلى نتيجة دالة إحصائياً (Palomares,1990). إنّ هذا الإجراء يحد من مشكلة عدم تحديد الحجم المطلوب لعينة الدراسة والتي عادة ما تحدّد بدون أي أساس منهجي أو علمي علماً بأنّ نتيجة الدلالة الإحصائية شديدة التأثر بحجم العينة، كما أنّ هذه الخطوة تجعل الباحث أكثر إدراكاً لأهمية ومفهوم حجم التأثير المطلوب في الدراسة (Thompson,1998d).

٢ - يجب حساب حجم التأثير (مؤشرات الدلالة العملية) بشكل مرافق لكل قيمة من قيم الاحتمالية المحسوبة ($r_{calculated}$) في البحث، بحيث لا تعرض أي نتيجة من نتائج اختبارات الدلالة الإحصائية بدون إرفاقها بمؤشر واحد على الأقل من مؤشرات حجم التأثير (McLean and Ernest,1998; Thompson and Snyder,1997b).

٣ - يجب عدم الاكتفاء بعرض نتائج مؤشرات حجوم التأثير المحسوبة فقط حسب الاستراتيجية السابقة، ولكن يجب أن تفسّر في جميع التحاليل سواء كانت مقاييس فروق معيارية أو نسب التباين المفسّر (Thompson,1994)، كما يجب إدراك أنّ تفسير نتائج حجوم التأثير والدلالة العملية هو أمر ذاتي Subjective ، حيث يعتمد بشكل كبير على أهداف البحث والدراسة، وذلك لكي لا تتكرّر

مشكلة مستويات الدلالة الإحصائية (٠,٠١ ، ٠,٠٥ ، ٠,٠٠١) التي قد تمنح الباحث شعوراً بالموضوعية، ولذا فإنّ العرف الذي وضعه كوهين في تصنيف وتفسير قيم حجم التأثير (صغير، متوسط، كبير) ليس مناسباً، حيث أنّ وجود قيمة لمؤشر حجم التأثير (d) أصغر من (٠,٢) والذي يصنفه كوهين بأنّه الحد الأدنى لحجم التأثير الصغير قد يعني تأثيراً كبيراً عند دراسة عقار على مرض عضال، فمثلاً النتيجة ذات حجم التأثير ٣٪ لأحد عقارات مرض الإيدز قد تكون معتبرة إذا كان الباحث أوالقارئ يمنح قيمة أعلى لحياة الإنسان، أو إذا كانت جميع الدراسات السابقة أنتجت حجم تأثير صفر٪، كل ما سبق يؤكد أنّ تفسير نتائج حجم التأثير أمر ذاتي ويجب على الباحث أن يؤديه باقتدار وألاّ يتخلى عن مسئوليته في تفسير النتائج أو أن يتعلّق بتقاليد علمية تبدو أنّها أكثر موضوعية لكنّها وللأسف قد تكون على حساب المعنى (Kellow,1998; Thompson,1994).

٤ - تفسير أثر حجم العينة في سياق تفسير نتائج اختبارات الدلالة الإحصائية. ويرى دانييل (١٩٩٨) بأنّ الباحث يجب عليه إجراء تحليل ما يسمى "ماذا لو What If" الذي يشير إلى الحد الأدنى من العينة الذي تكون فيه العينات المختلفة عند ثبات التأثير تصبح دائماً دالة إحصائياً، وكذلك الحد الأدنى من حجم العينة الذي تكون فيه نتائج العينات المختلفة غير دالة إحصائياً (Daniel,1998a; Daniel,1997)، وهذا يعني أنّه يجب على الباحث عند تقويم نتائج اختبار الدلالة الإحصائية أن يفسّر أثر حجم العينة (Palomares,1990). ولقد عرض سبينسر Spencer(1995) جدولاً يوضح مثالاً مبسطاً لأسلوب تفسير أثر حجم العينة على نتائج التحاليل الإحصائية.

جدول ١٢ : نموذج من تفسير أثر حجم العينة في سياق تفسير النتائج

نتيجة الدلالة الإحصائية	حجم العينة	الاستنتاج
دالة إحصائياً	صغير	نتيجة مهمة
دالة إحصائياً	كبير	ربما تكون النتيجة لها أهمية عملية وربما لا
غير دالة إحصائياً	صغير	لا يمكن الاستنتاج
غير دالة إحصائياً	كبير	ربما تكون الفرضية الصفرية خاطئة فعلاً

٥ - ويرى ماكلين وإرنست McLean and Ernest(1998) بأنّ على الباحث أن يلاحظ أنّ أهمية استخدام اختبارات الدلالة الإحصائية تقل كلما زاد حجم العينة، بينما تزداد في تلك الحالة أهمية الاعتماد على حسابات الدلالة العملية، وهذا يعني أنّ على الباحث أن يمنح مقاييس الدلالة العملية وزناً أكبر في تفسير النتائج كلما زاد حجم العينة، بينما تزداد أهمية نتائج اختبارات الدلالة الإحصائية في تفسير النتائج كلما قل حجم العينة وذلك لأنّ قيمة الاحتمالية المحسوبة من اختبارات الدلالة الإحصائية هي التي تساعدنا في توضيح البيانات المحصلة وما إذا كان سببها الصدفة أم لا.

٦ - استخدام تقديرات حدود الثقة التي يمكن أن تكون مفيدة في تفسير نتائج الدلالة الإحصائية، حيث أنّها تزود الباحث بفرصة أفضل لتفسير الحدود الدنيا والحدود العليا لاحتمالية نتائج العينة، كما أنّها تساعد الباحث في تفسير أثر حجم العينة خاصة في دقة القياس (McLean and Ernest,1998)، وفوائد أخرى سبقت مناقشتها في هذا الفصل.

٧ - إذا كانت نتائج الدراسة غير دالة إحصائياً يفترض أن يقوم الباحث بتفسير تلك النتيجة وعدم الاكتفاء بالإشارة إلى عدم دلالتها إحصائياً، كما يفترض أن يقوم الباحث بعرض نتيجة اختبار القوة الإحصائية وذلك للتأكد من أنّ الاختبار الإحصائي المنفذ لا يعاني من انخفاض القوة، وأنّ عدم الحصول على الدلالة الإحصائية لم يكن بسبب انخفاض القوة الإحصائية للاختبار المستخدم (Daniel,1998a; Daniel,1997).

٨ - مع أهمية الاستراتيجيات السابقة إلاّ أنّ أهم ممّا سبق هو عدم اعتبار الباحث أنّ الفرضية أو النظرية مبرهنة ومثبتة حتى لو ثبتت دلالتها الإحصائية ودلالتها العملية حتى يثبت للباحث أنّ النتائج قابلة للإعادة، بدون تحاليل الإعادة لن يستطيع الباحث أن يستنتج احتمالية إعادة تلك النتائج في عينات مستقبلية أو في المجتمع، تحليل الإعادة فقط هو الذي يزوّد الباحث بهذا الاستنتاج (McLean and Ernest,1998).

٩ - إذا لم يتمكن الباحث من تنفيذ الإعادة الخارجية لدراسة ما، فإنّ أساليب تحليل الإعادة الداخلية مثل Cross-validation, Bootstrap, and Jackknife ربما يكون مناسباً (McLean and Ernest,1998; Thompson and Snyder,1997a; Palomares,1990; Thpmpson,1994).

ثانياً

إجابة السؤالين الرابع والخامس

تمهيد

قام الباحث من خلال الجزء الأول من هذا الفصل بالإجابة عن التساؤلات الثلاث الأولى الرئيسة في هذه الدراسة وهي:

السؤال الأول: ما أهم مشكلات الدلالة الإحصائية في البحث التربوي؟

السؤال الثاني: ما أهم المفاهيم الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في البحث التربوي؟

السؤال الثالث: ما الأساليب الجديدة لحل مشكلات الدلالة الإحصائية في البحث التربوي؟

وسيقوم الباحث في الجزء الثاني من هذا الفصل بالإجابة عن التساؤلين الرابع والخامس وهما على النحو التالي:

السؤال الرابع: ما واقع مشكلات الدلالة الإحصائية في رسائل درجة الماجستير بكلية التربية بجامعة أم القرى بمكة المكرمة؟

السؤال الخامس: ما مدى وجود المفاهيم الإحصائية التي يمكن أن تقدّم حلولاً لمشكلات الدلالة الإحصائية في رسائل درجة الماجستير بكلية التربية بجامعة أم القرى بمكة المكرمة؟

ومن خلال ما توصل إليه الباحث في الجزء الأول من هذا الفصل من تحديد لمشكلات الدلالة الإحصائية في البحث التربوي، وأهم المفاهيم والأساليب الإحصائية التي يمكن أن تقدّم حلولاً لتلك المشكلات، قام الباحث بتحديد عدّة تساؤلات فرعية للسؤالين الرئيسين الرابع والخامس على النحو التالي:

التساؤلات الفرعية للسؤال الرابع

للإجابة على السؤال الرابع (ما واقع مشكلات الدلالة الإحصائية في الرسائل العلمية بكلية التربية بجامعة أم القرى بمكة المكرمة؟) حدّد الباحث التساؤلات الفرعية التالية:

(أ) ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى نتائج الدلالة الإحصائية كتفسير لأهمية النتائج؟

(ب) ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى نتائج اختبارات الدلالة الإحصائية كتفسير لتأثير المعالجة أو المتغيّرات المستقلة؟

(ج) ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى قيم الاحتمالية (P) لتقدير حجم التأثير؟

(د) ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى قيم الاحتمالية (r) كتفسير لاحتمالية الفرضية الصفرية نفسها؟

(ه) ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى قيم الاحتمالية (r) لتقدير احتمالية إعادة النتائج؟

(و) ما مدى تحيّز رسائل الماجستير في كلية التربية بجامعة أم القرى للنتائج الدالة إحصائياً؟

(ز) ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى للتساؤلات البحثية عن الفروق ذات الدلالة الإحصائية؟

(ح) ما مدى وجود نتائج معبّرة عن مأزق الطبيعة الثنائية لنتائج اختبارات الدلالة الإحصائية في رسائل الماجستير بكلية التربية بجامعة أم القرى؟

التساؤلات الفرعية للسؤال الخامس

وللإجابة عن السؤال الخامس (ما مدى وجود المفاهيم الإحصائية التي يمكن أن تقدّم حلولاً لمشكلات الدلالة الإحصائية في رسائل درجة الماجستير بكلية التربية بجامعة أم القرى بمكة المكرمة؟) حدّد الباحث التساؤلات الفرعية التالية:

(٥ - ١) ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى لتحليل حدود الثقة في تفسير نتائج اختبارات الدلالة الإحصائية؟

(٥ - ٢) ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى تحليل القوة الإحصائية في تفسير نتائج اختبارات الدلالة الإحصائية؟

(٥ - ٣) ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى لمقاييس حجم التأثير لتفسير نتائج اختبارات الدلالة الإحصائية؟

(٥ - ٤) ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى لتقديرات إعادة النتائج؟

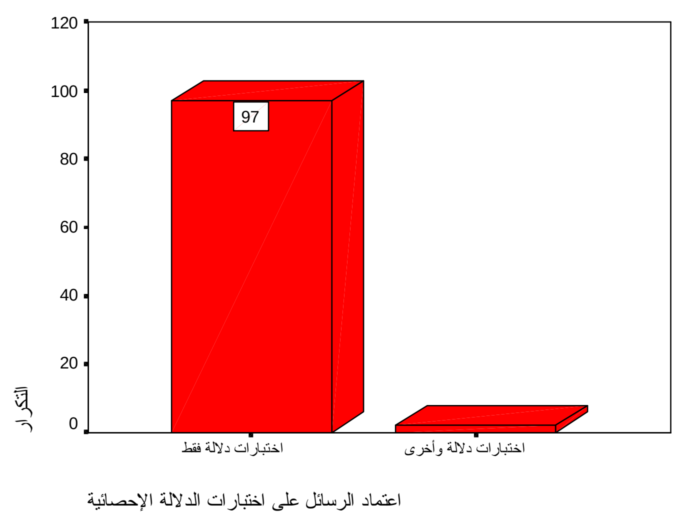
إجابة السؤال الرابع

إجابة السؤال (٤ – ١): ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى لنتائج اختبارات الدلالة الإحصائية كتفسير لأهمية النتائج؟

للإجابة عن هذا التساؤل قام الباحث بحصر الرسائل التي اكتفت بعرض نتائج اختبارات الدلالة الإحصائية واعتبرتها نهاية البحث وغايته دون إجراء أي تحليلات أخرى، كما هو موضح بالجدول (١٣) والرسم البياني (١٤) التاليين:

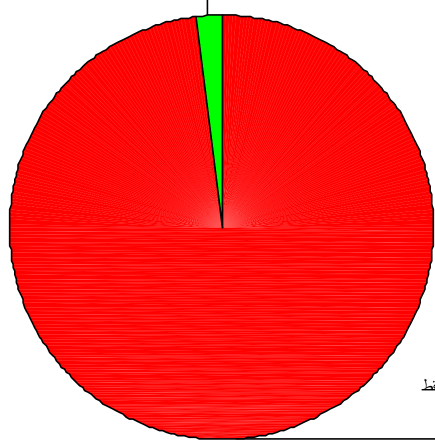
جدول ١٣ : التكرارات والنسب المئوية للرسائل التي اعتمدت على نتائج الدلالة الإحصائية واعتبرتها منتهى الدراسة

الحـــالـــه	التكرارات	النسبة المئوية
الرسائل المعتمدة على اختبارات الدلالة الإحصائية فقط	٩٧	٩٨,٠٪
الرسائل التي لم تعتمد فقط على اختبارات الدلالة الإحصائية	٢	٢,٠٪
المجموع	٩٩	١٠٠٪



رسم بياني ١٤ : تكرارات الرسائل التي اعتمدت على نتائج اختبارات الدلالة الإحصائية والتي اعتبرتها منتهى الدراسة

يتضح من الجدول والرسم البياني السابقين أنّ عدد (٩٧) رسالة من رسائل الماجستير في عينة الدراسة اعتمدت على نتائج اختبارات الدلالة الإحصائية فقط في تفسير النتائج، حيث اعتبرت نتائج تلك الاختبارات غاية الدراسة ومنتهاها، بينما بلغ عدد الرسائل التي لم تعتمد فقط على نتائج اختبارات الدلالة الإحصائية واستخدمت أساليب أخرى رسالتين فقط، حيث استخدمتا مقاييس للدلالة العملية.



رسم بياني ١٥ : النسبة المئوية للرسائل التي اعتمدت على نتائج الدلالة الإحصائية والتي اعتبرتها منتهى الدراسة

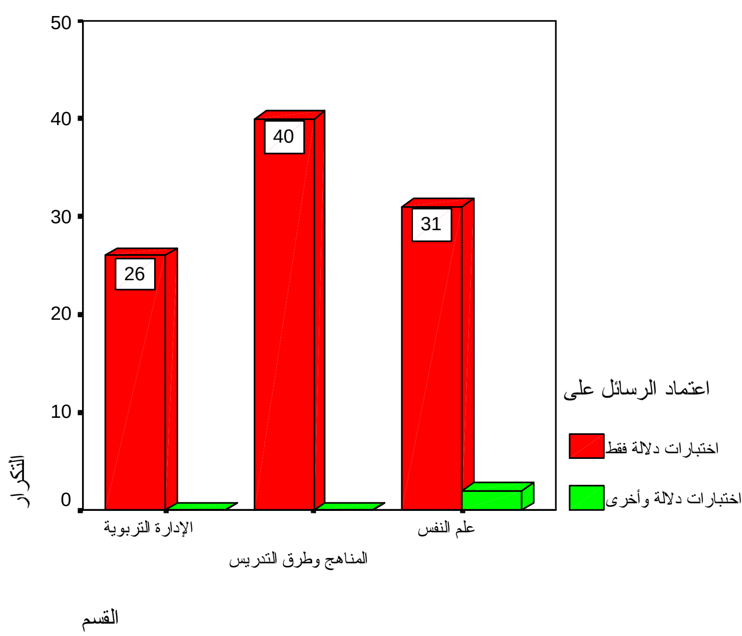
الرسم البياني السابق يوضح نسبة عدد الرسائل العلمية التي اعتمدت فقط في نتائجها على نتائج اختبارات الدلالة الإحصائية (٩٨٪) مقارنة بـ (٢٪) التي لم تعتمد على نتائج اختبارات الدلالة الإحصائية فقط في تحليل وتفسير نتائج الدراسة، بل اعتمدت أيضاً على اختبارات الدلالة العملية.

وبمقارنة هذه النتيجة مع نتائج دراسات سابقة يتضح أنّ هذه المشكلة أكثر حدّة في رسائل الماجستير بكلية التربية بجامعة أم القرى، حيث أنّ نسب الدراسات التي اعتمدت فقط على نتائج اختبارات الدلالة الإحصائية في تفسير أهمية النتائج في الدراسات السابقة كانت في أعلى تقدير لها (٨٧٪) كما هي في دراسة ثومبسون Thompson(1999)، بينما كانت النسبة في دراسة ثومبسون وسنايدر Thompson & Snyder(1997) هي (٣٦٪)، أمّا في دراسة كيرك Kirk(1996) فقد بلغت تلك النسبة (٢٣٪).

الجدول والرسم البياني التاليان يوضحان الفروق بين الأقسام الثلاث في كلية التربية في استخدام نتائج الدلالة الإحصائية كتفسير لأهمية النتائج:

جدول ١٤: تكرارات الرسائل التي اعتمدت على نتائج اختبارات الدلالة الإحصائية فقط حسب أقسام كلية التربية

الحـــالـــه	الإدارة	المناهج	علم النفس
الرسائل المعتمدة على اختبارات الدلالة الإحصائية فقط	٢٦	٤٠	٣١
الرسائل التي لم تعتمد فقط على اختبارات الدلالة الإحصائية	٠	٠	٢
المجموع	٢٦	٤٠	٣٣



رسم بياني ١٦: تكرارات الرسائل التي اعتمدت على نتائج اختبارات الدلالة الإحصائية فقط حسب أقسام كلية التربية

يتضح من الجدول والرسم البياني السابقين أنّ رسالتين فقط في عينة الدراسة لم تعتمدا على نتائج اختبارات الدلالة الإحصائية كنتائج وحيدة تعكس أهمية الدراسة، بل تابعتا اختبارات الدلالة الإحصائية باختبارات الدلالة العملية، وهاتين الرسالتين تخصان قسم علم النفس. ولعلّ سبب وجود تلك الرسائل التي لم تعتمد فقط على اختبارات الدلالة

الإحصائية في نتائج دراستها بل أجرت اختبارات إحصائية أخرى مكملة لاختبارات الدلالة الإحصائية في قسم علم النفس هو وجود تخصص دقيق بالقسم يعنى بالإحصاء والبحوث.

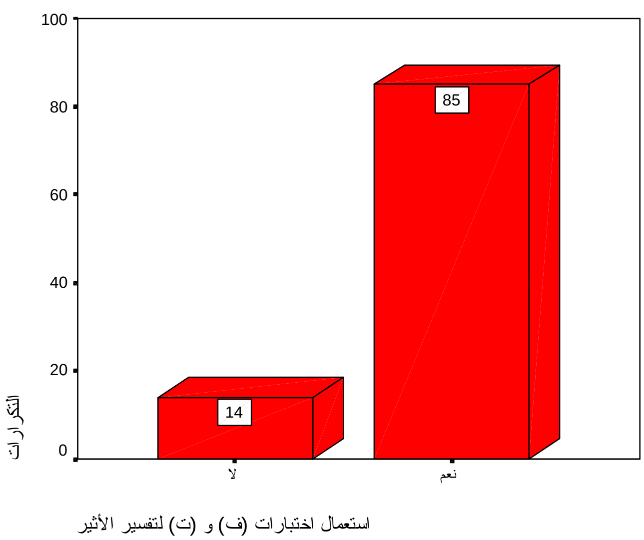
إجابة السؤال (٤ – ٢): ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى لنتائج اختبارات الدلالة الإحصائية كتفسير لتأثير المعالجات أو المتغيّرات المستقلة؟

للإجابة عن هذا التساؤل قام الباحث بحساب تكرارات اختبارات الفروق بين المتوسطات باستخدام اختبار (t) واختبارات تحليل التباين (F) التي استخدمها طلاب الماجستير في رسائلهم بغرض معرفة أثر المعالجات أو المتغيّرات المستقلة على المتغيّرات التابعة، وبدون استخدام أي مؤشرات أخرى.

الجدول والرسم البياني التاليان يوضحان المشاهدات التالية:

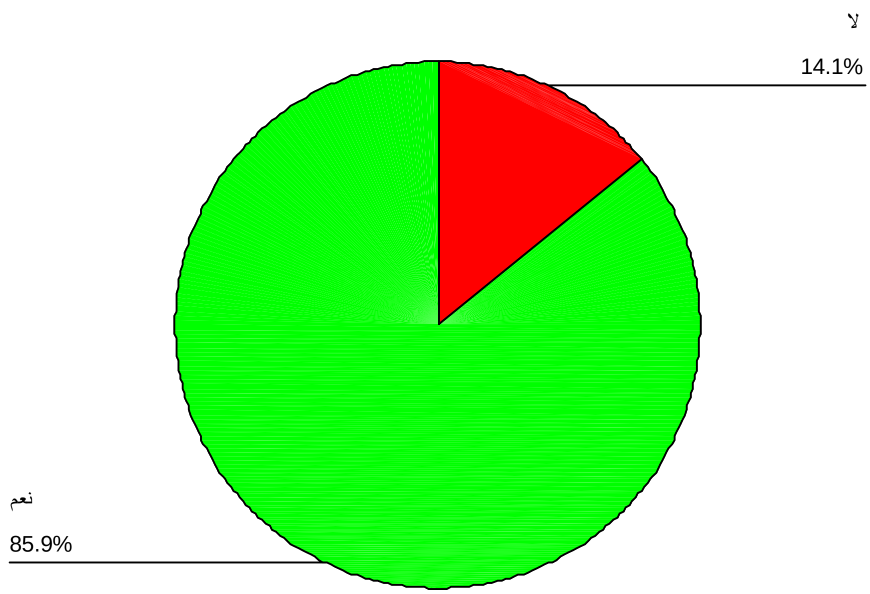
جدول ١٥ : تكرارات اختبارات **t** و **F** التي استخدمت لتفسير تأثير المعالجات أو المتغيرات المستقلة

الحـــالـــه	التكرارات	النسبة المئوية
الرسائل المستخدمة لاختبارات t و F لتفسير التأثيرات	٨٥	٪٨٥,٩
الرسائل غير المستخدمة لاختبارات t و F لتفسير التأثيرات	١٤	٪١٤,١
المجموع	٩٩	٪١٠٠



رسم بياني ١٧ : تكرارات اختبارات **t** و **F** التي استخدمت لتفسير تأثير المعالجات أو المتغيرات المستقلة

يتضح من الجدول والرسم البياني السابقين أنّ عدد الرسائل العلمية التي استخدمت اختبارات الدلالة الإحصائية كتفسير لتأثير المعالجة أو المتغيّرات المستقلة (٨٥) رسالة مقابل (١٤) رسالة لم تستخدم اختبارات الدلالة الإحصائية كتفسير لتأثير المعالجة أو المتغيّرات المستقلة.



رسم بياني ١٨ : النسبة المئوية لاختبارات t و F التي استخدمت لتفسير تأثير المعالجات أو المتغيرات المستقلة

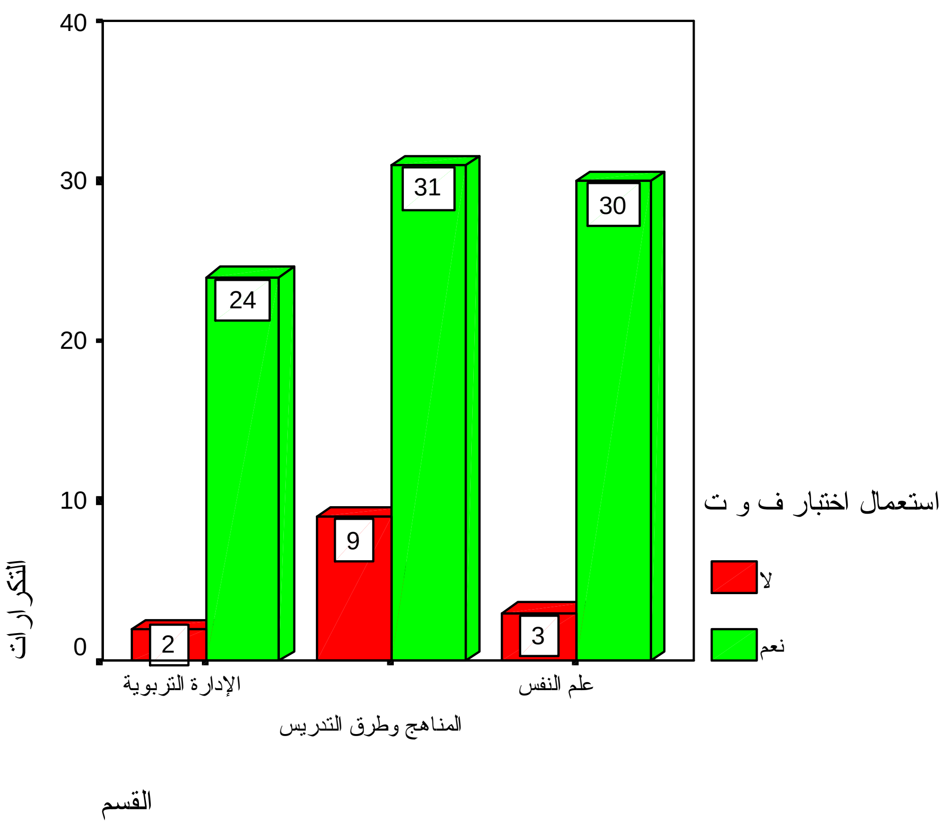
الرسم البياني السابق يوضح نسبة رسائل الماجستير التي استخدمت نتائج اختبارات الدلالة الإحصائية كتفسير لتأثير المعالجات أو المتغيرات المستقلة حيث بلغت نسبتها (٨٥,٩٪).

تقترب هذه النتيجة مع النتيجة التي توصلت إليها دراسة ماكلين(١٩٩٥) حيث وجدت أنّ نسبة الدراسات التي اعتمدت على اختبارات (t) و (F) في تفسير التأثير والتي لم ترفق بمؤشرات لحجم التأثير قد بلغت (٩٣٪)، علماً بأنّ النسبة في رسائل الماجستير بكلية التربية بجامعة أم القرى كانت في وضع أفضل حيث بلغت (٨٥,٩٪).

الجدول والشكل البياني التاليان يوضحان الفروق بين الأقسام الثلاثة في كلية التربية في استخدام نتائج اختبارات الدلالة الإحصائية كتفسير لأثر المعالجة أو المتغيّرات المستقلة:

جدول ١٦ : تكرارات اختبارات (t) و (F) التي استخدمت لتفسير تأثير المعالجات أو المتغيرات المستقلة حسب أقسام كلية التربية

الحـــالـــة	الإدارة التربوية		المناهج		علم النفس	
	ت	٪	ت	٪	ت	٪
الرسائل المستخدمة لاختبارات t و F لتفسير التأثيرات	٢٤	٩٢,٣	٣١	٧٧,٥	٣٠	٩٠,٩
الرسائل غير المستخدمة لاختبارات t و F لتفسير التأثيرات	٢	٧,٧	٩	٢٢,٥	٣	٩,١
المجموع	٢٦	١٠٠	٤٠	١٠٠	٣٣	١٠٠



رسم بياني ١٩ : تكرارات اختبارات t و F التي استخدمت لتفسير تأثير المعالجات أو المتغيرات المستقلة حسب أقسام كلية التربية

من الرسم البياني السابق يتضح أنّ أعلى نسبة لاستخدام اختبارات الدلالة الإحصائية في تفسير أثر المعالجة أو المتغيّرات المستقلة هو في قسم الإدارة التربوية والتخطيط (٩٢,٣٪)، ثمّ في قسم علم النفس (٩٠,٩٪)، وأخيراً قسم المناهج وطرق التدريس (٧٧,٥٪).

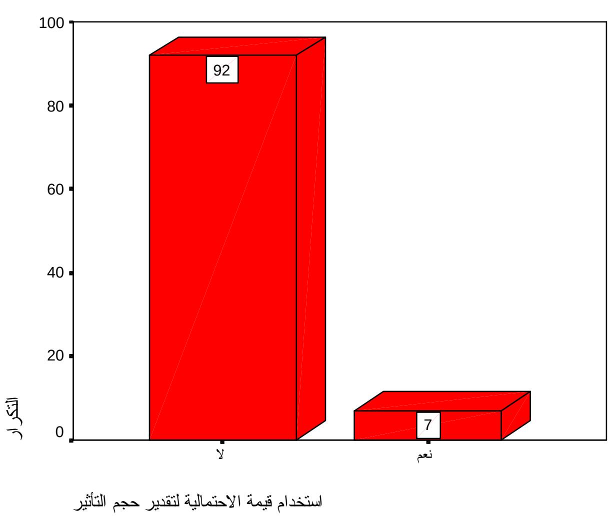
إجابة السؤال (٤ – ٣): ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى قيمة الاحتمالية (r) كتقدير لحجم التأثير؟

للإجابة عن هذا التساؤل قام الباحث بحساب تكرارات الرسائل التي استخدمت عبارات وصف لقيمة الاحتمالية (r) مثل عبارات (عالية الدلالة ، ذات دلالة إحصائية عالية جداً ، مرتفعة الدلالة ، ... وهكذا). الجدول (١٧) يوضح ما يلي:

جدول ١٧ : تكرارات الرسائل التي استخدمت عبارات لوصف قيمة الاحتمالية (**p**)

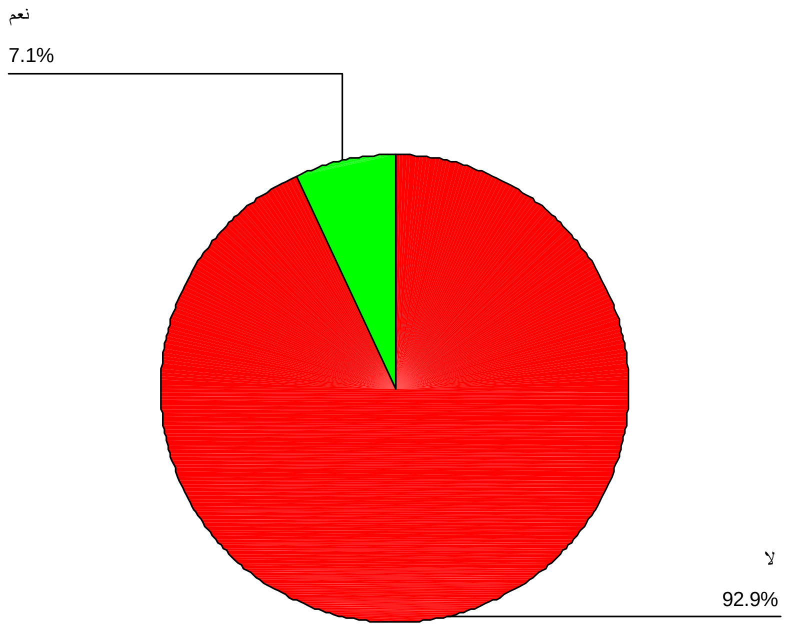
الحـــالـــه	التكرارات	النسبة المئوية
الرسائل التي استخدمت قيمة (r) لتقدير حجم التأثير	٧	٪ ٧,١
الرسائل التي لم تستخدم قيمة (r) لتقدير حجم التأثير	٩٢	٪٩٢,٩
المجموع	٩٩	٪١٠٠

الرسم البياني التالي يوضح تكرارات الرسائل التي استخدمت عبارات لوصف قيمة الاحتمالية المحسوبة (r) .



رسم بياني ٢٠ : تكرارات الرسائل التي استخدمت قيمة الاحتمالية المحسوبة (**p**) لتقدير حجم التأثير

الرسم البياني التالي يوضح النسبة المئوية لتكرارات الرسائل التي استخدمت عبارات لوصف قيمة الاحتمالية المحسوبة (r) .



رسم بياني ٢١ : النسبة المئوية لتكرارات الرسائل التي استخدمت قيمة الاحتمالية المحسوبة (**p**) لتقدير حجم التأثير

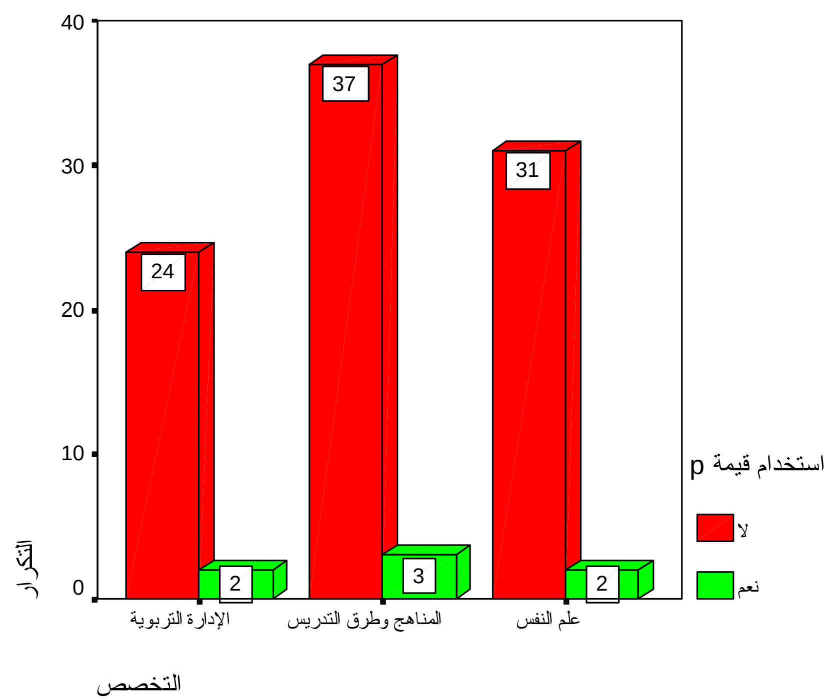
يتضح من الجدول السابق وكذلك الرسمين البيانيين السابقين أنّ الرسائل العلمية وبنسبة (٩٢,٩٪) لم تستخدم عبارات لتقويم قيمة الاحتمالية (r) كقيمة لحجم التأثير، وهذا يعني أنّ هذه المشكلة موجودة في رسائل الماجستير بكلية التربية بجامعة أم القرى بمكة المكرمة بدرجة بسيطة. وتتفق نتيجة هذه الدراسة عمّا وجده كل من بلوكير Plucker(1997) وهاركوم Harcum(1989) من وجود لعبارات مثل (عالية الدلالة بوضوح جداً في عينتي دراستيهما.

الجدول التالي يوضح تكرارات الرسائل التي استخدمت قيمة الاحتمالية المحسوبة (r) لتقدير حجم التأثير:

جدول ١٨ : تكرارات الرسائل التي استخدمت قيمة (**p**) لتقدير حجم التأثير حسب أقسام كلية التربية

الحـــالـــه	الإدارة التربوية	المناهج	علم النفس
تكرارات الرسائل التي استخدمت قيمة (*r*) لتقدير حجم التأثير	٢	٣	٢
تكرارات الرسائل التي لم تستخدم قيمة (*r*) لتقدير حجم التأثير	٢٤	٣٧	٣١

الرسم البياني التالي يوضح تمثيلاً لتكرارات الرسائل التي استخدمت قيمة الاحتمالية المحسوبة (*r*) لتقدير حجم التأثير:



رسم بياني ٢٢ : تكرارات الرسائل التي استخدمت قيمة الاحتمالية المحسوبة (**p**) لتقدير حجم التأثير حسب الأقسام

يتضح من الرسم البياني السابق أنّ عدد تكرارات الرسائل التي استخدمت قيمة الاحتمالية المحسوبة (*r*) كتقدير لحجم التأثير متساوية تقريباً في جميع أقسام كلية التربية بمكة المكرمة.

إجابة السؤال (٤ – ٤): ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى لقيمة الاحتمالية (r) كتفسير لاحتمالية الفرضية الصفرية نفسها؟

للإجابة عن هذا التساؤل قام الباحث بحساب تكرارات الرسائل التي استخدمت قيم الاحتمالية (r) لتفسير احتمالية الفرضية الصفرية نفسها. الجدول التالي تلك التكرارات على النحو التالي:

جدول ١٩ : تكرارات الرسائل التي استخدمت قيمة الاحتمالية (**p**) لتفسير احتمالية الفرضية الصفرية نفسها

الحـــالـــه	التكرارات	النسبة المئوية
الرسائل التي استخدمت قيمة (r) لتفسير احتمالية H_o	٠	صفر ٪
الرسائل التي لم تستخدم قيمة (r) لتفسير احتمالية H_o	٩٩	١٠٠٪
المجموع	٩٩	١٠٠٪

يتضح من الجدول السابق أنّ عدد الرسائل التي استخدمت قيمة (r) لتفسير احتمالية الفرضية الصفرية نفسها هي صفر، وهذا يعني أنّ هذه المشكلة غير موجودة في رسائل الماجستير بكلية التربية بجامعة أم القرى.

إجابة السؤال (٤ – ٥): ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى لقيمة الاحتمالية (r) كتقدير لاحتمالية إعادة النتائج؟

للإجابة عن هذا التساؤل قام الباحث بحصر تكرارات الرسائل التي فسرت قيمة (r) كمؤشر لتقدير إعادة النتيجة. الجدول التالي يبيّن النتيجة كما يأتي:

جدول ٢٠ : تكرارات الرسائل التي ورد فيها تفسير لقيمة (**p**) كمؤشر لإعادة النتيجة

الحـــالـــه	التكرارات	النسبة المئوية
الرسائل التي استخدمت قيمة (r) كمؤشر لإعادة النتيجة	٠	صفر ٪
الرسائل التي لم تستخدم قيمة (r) كمؤشر لإعادة النتيجة	٩٩	١٠٠٪
المجموع	٩٩	١٠٠٪

يتضح من الجدول السابق أنّ عدد تكرارات تفسير قيمة (r) لمؤشر لاحتمالية إعادة النتائج هو (صفر)، أي أنّ هذه المشكلة ليس لها وجود فعلي في رسائل الماجستير بكلية التربية بجامعة أم القرى.

إجابة السؤال (٤ – ٦): ما مدى تحيّز رسائل الماجستير في كلية التربية بجامعة أم القرى للنتائج الدالة إحصائياً؟

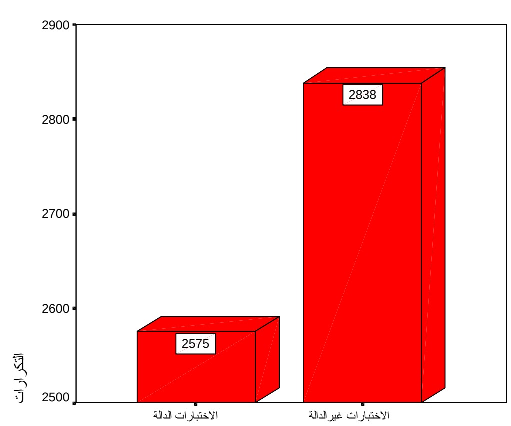
للإجابة عن هذا التساؤل قام الباحث بحساب مجموع تكرارات النتائج الدالة إحصائياً ومجموع تكرارات النتائج غير الدالة إحصائياً. الجدول(٢١) يوضح هذه التكرارات ومتوسطاتها على النحو التالي :

جدول ٢١ : مجموع تكرارات النتائج الدالة وغير الدالة إحصائياً ومتوسطاتها

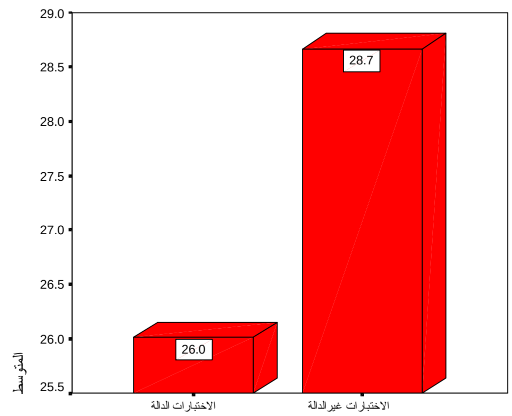
الحـــالـــه	التكرارات	المتوسط لكل رسالة
النتائج التي وصلت إلى مستوى الدلالة الإحصائية	٢٥٧٦	٢٦,٠١
النتائج التي لم تصل إلى مستوى الدلالة الإحصائية	٢٨٣٨	٢٨,٦٧
المجموع	٤١٤٥	

يلاحظ من الجدول السابق أنّ مجموع تكرارات النتائج الدالة إحصائياً بلغ (٢٥٧٥) نتيجة وبمتوسط حسابي (٢٦,٠١) نتيجة، مقارنة بمجموع تكرارات النتائج غير الدالة إحصائياً الذي بلغ (٢٨٣٨) نتيجة وبمتوسط حسـابي قدره (٢٨,٦٧) نتيجة.

الرسمان البيانيان التاليان يمثّلان الفرق بين مجموع النتائج الدالة إحصائياً والنتائج غير الدالة إحصائياً، وكذلك الفرق بين متوسط النتائج الدالة إحصائياً والنتائج غير الدالة إحصائياً على التوالي:

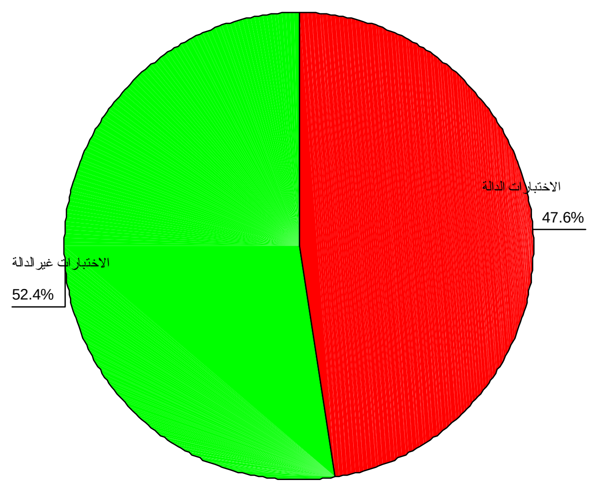


رسم بياني ٢٣ : مجموع تكرارات النتائج الدالة إحصائياً والنتائج غير الدالة إحصائياً



رسم بياني ٢٤ : متوسط تكرارات النتائج الدالة إحصائياً والنتائج غير الدالة إحصائياً

الرسم البياني التالي يوضح النسبة المئوية لإجمالي الاختبارات الدالة إحصائياً والاختبارات غير الدالة إحصائياً.



رسم بياني ٢٥ : النسبة المئوية لتكرارات النتائج الدالة إحصائياً والنتائج غير الدالة إحصائياً

يتضح من الرسم البياني السابق أنّ نسبة النتائج غير الدالة إحصائياً هي (٥٢,٤٪) ، مقارنة بنسبة النتائج الدالة إحصائياً وهي (٤٧,٦٪).

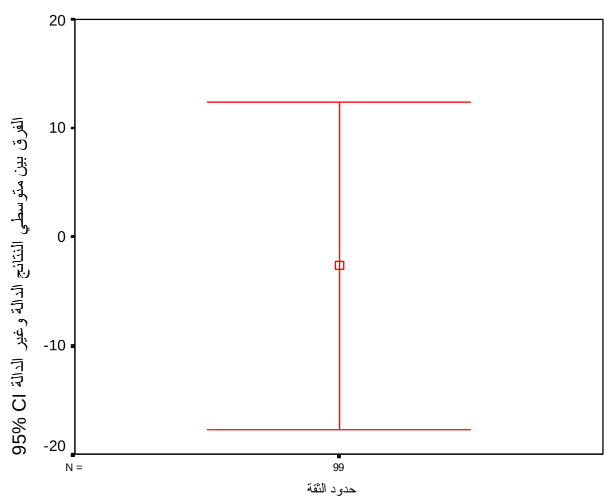
تشير النتائج السابقة أنّه لا يوجد تحيّز من قبل رسائل الماجستير في كلية التربية بجامعة أم القرى للنتائج التي وصلت إلى مستوى الدلالة الإحصائية مقابل النتائج التي لم تصل إلى مستوى الدلالة الإحصائية، فلم تكن الفروق في صالح النتائج الدالة إحصائياً ولو بفارق قليل، بل على العكس تماماً كانت الفروق في صالح النتائج غير الدالة إحصائياً، بمعنى أنّه لو وجد تحيّز في نتائج اختبارات الدلالة الإحصائية في رسائل الماجستير بكلية التربية بجامعة أم القرى فقد يكون هذا التحيّز لصالح النتائج غير الدالة إحصائياً.

الجدول التالي يوضح تقدير حدود الثقة عند مستوى (٩٥٪) للفرق بين متوسطي النتائج الدالة إحصائياً والنتائج غير الدالة إحصائياً على النحو التالي:

جدول ٢٢ : تقدير حدود الثقة عند مستوى ٩٥٪ للفرق بين متوسطي النتائج الدالة إحصائياً وغير الدالة إحصائياً

الحـــالـــه	الفرق بين المتوسطات	الحد الأعلى	الحد الأدنى
حدود الثقة عند مستوى ٩٥٪ للفرق بين متوسطي النتائج الدالة إحصائياً والنتائج غير الدالة إحصائياً	-٢,٦٦	-١٧,٧٢	١٢,٤١

كما أنّ الرسم البياني التالي لأعمدة الخطأ يوضح تمثيل حدود الثقة عند مستوى ٩٥٪ لتقدير الفرق بين متوسطي النتائج الدالة إحصائياً والنتائج غير الدالة إحصائياً على النحو التالي:



رسم بياني ٢٦ : تمثيل حدود الثقة لمتوسطي النتائج الدالة إحصائياً وغير الدالة إحصائياً

يتضح من الجدول السابق أنّ الفرق بين متوسطي العينتين غير المستقلتين والذي بلغ (٢,٦٦) لصالح النتائج غير الدالة إحصائياً لها حدود ثقة عند مستوى ٩٥٪ مقداره (من –١٧,٧٢ إلى ١٢,٤١) ، وحيث أنّ قيمتي حدود الثقة تتضمنان (الصفر) فإنّ هذا يعني عدم وجود فروق ذات دلالة إحصائية بين متوسط النتائج الدالة إحصائياً والنتائج غير الدالة إحصائياً، ويوضح الرسم البياني (٢٣) السابق أنّ حدود الثقة عند مستوى ٩٥٪ للفرق بين متوسطي العينتين تتضمن الصفر، ممّا يعني عدم وجود فروق ذات دلالة إحصائية بينهما.

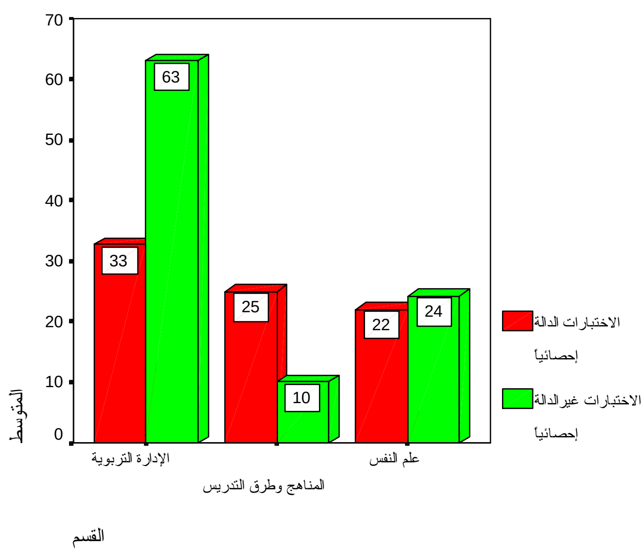
هذه النتيجة التي توصلت إليها هذه الدراسة لا تتفق مع ما قرّره كل من ويلكنسون Wilkinson(1992)، وكيستر Keaster(1988)، ودانييل Daniel(1997)، وهوستون Huston(1993)، ونيكس وبارنيت Nix & Barnette(1998) الذين أشاروا إلى وجود تحيّز من قبل الباحثين للنتائج التي وصلت إلى مستوى الدلالة الإحصائية مقابل النتائج التي لم تصل إلى مستوى الدلالة الإحصائية.

الجدول التالي يوضح متوسطات عدد النتائج الدالة إحصائياً والنتائج غير الدالة إحصائياً حسب الأقسام:

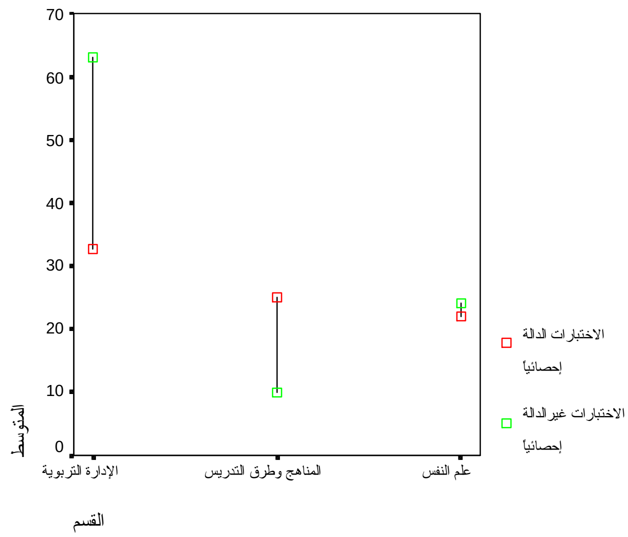
جدول ٢٣ : مقارنة متوسطات عدد النتائج الدالة إحصائياً وغير الدالة إحصائياً حسب أقسام كلية التربية

الحـــالـــه	الإدارة التربوية	المناهج وطرق التدريس	علم النفس
متوسط عدد النتائج الدالة إحصائياً	٣٢,٦٩	٢٤,٩٥	٢٢,٠٣
متوسط عدد النتائج غير الدالة إحصائياً	٦٣,٠٤	١٠,٠	٢٤,٢

الرسمان البيانيان التاليان يمثلان متوسطات النتائج الدالة إحصائياً وغير الدالة إحصائياً حسب أقسام كلية التربية:



رسم بياني ٢٧: تمثيل حدود الثقة لمتوسط النتائج الدالة إحصائياً وغير الدالة إحصائياً في كل قسم من أقسام كلية التربية



رسم بياني ٢٨ : تمثيل الفرق بين متوسطات النتائج الدالة إحصائياً وغير الدالة إحصائياً حسب أقسام كلية التربية

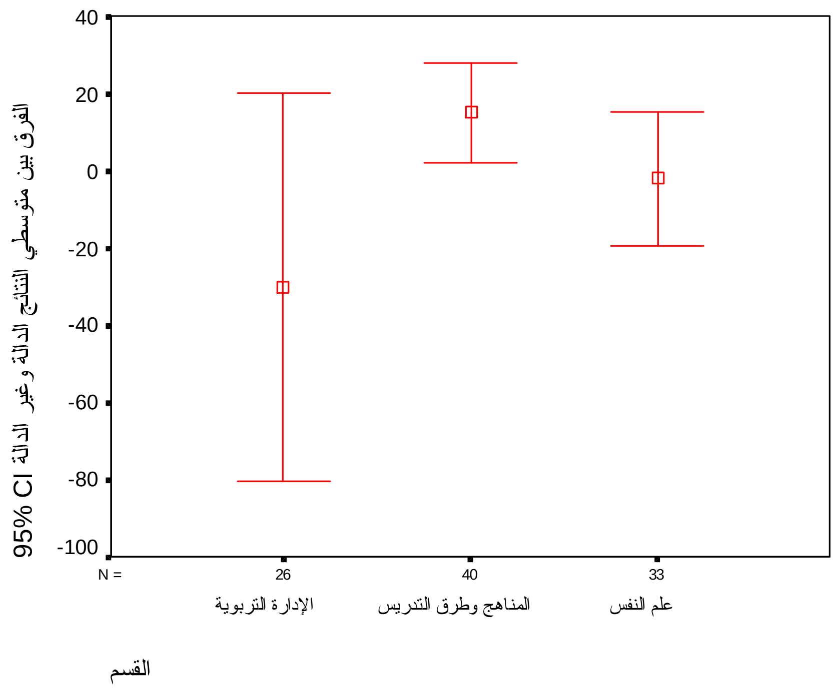
يتضح من الرسمين البيانيين السابقين أنّ أقل فروق بين متوسطي النتائج الدالة إحصائياً وغير الدالة إحصائياً هي في قسم علم النفس وهي في ذلك القسم لصالح النتائج غير الدالة إحصائياً، أمّا أعلى فرق بين متوسطي النتائج الدالة إحصائياً والنتائج غير الدالة إحصائياً فهي في قسم الإدارة التربوية والتخطيط وهي في صالح النتائج غير الدالة إحصائياً، وأمّا قسم المناهج وطرق التدريس فهو وسط بين القسمين السابقين من حيث الفرق بين متوسطي النتائج الدالة إحصائياً والنتائج غير الدالة إحصائياً، والفرق في صالح النتائج الدالة إحصائياً.

الجدول التالي يوضح الفروق بين أقسام كلية التربية في درجة التحيّز للنتائج غير الدالة إحصائياً:

جدول ٢٤ : الفروق في درجة التحيّز للنتائج غير الدالة إحصائياً حسب أقسام كلية التربية

الحـــالـــة	الإدارة التربوية	المناهج وطرق التدريس	علم النفس
الفرق بين متوسطي النتائج الدالة وغير الدالة إحصائياً	-٣٠,٣٥	١٤,٩٥	-٢,١٨
الحد الأعلى لحدود الثقة عند مستوى ٩٥٪	١٩,٧٦	٢٧,٩٠	١٥,٣٥
الحد الأدنى لحدود الثقة عند مستوى ٩٥٪	-٨٠,٤٥	٢,٠٠	-١٩,٧١

الرسم البياني التالي يمثل حدود الثقة عند مستوى ٩٥٪ للفروق بين متوسطات النتائج الدالة إحصائياً والنتائج غير الدالة إحصائياً في كل قسم من أقسام كلية التربية:



رسم بياني ٢٩ : تمثيل حدود الثقة لمتوسط النتائج الدالة إحصائياً وغير الدالة إحصائياً في كل قسم من أقسام كلية التربية

يتضح من الرسم البياني السابق أنّ هناك فروقاً ذات دلالة إحصائية بين متوسط النتائج غير الدالة إحصائياً والنتائج الدالة إحصائياً في قسم المناهج وطرق التدريس فقط حيث أنّ حدود الثقة للفرق بين متوسط النتائج الدالة إحصائياً وغير الدالة إحصائياً لا تتضمن الصفر. أمّا في قسمي الإدارة التربوية والتخطيط وعلم النفس فإنّ الفروق بين متوسطي النتائج الدالة إحصائياً والنتائج غير الدالة إحصائياً ليست ذات دلالة إحصائية بسبب أنّ حدود الثقة فيهما تتضمنان الصفر.

إجابة السؤال (٤ – ٧): ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى للتساؤلات البحثية عن الفروق ذات الدلالة الإحصائية؟

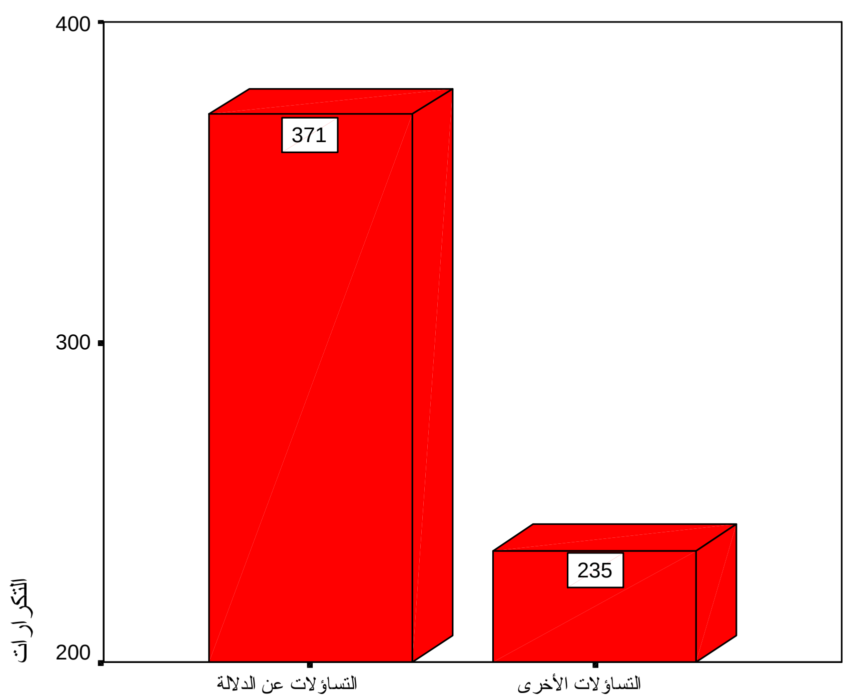
للإجابة عن هذا التساؤل قام الباحث بحساب مجموع التساؤلات عن الفروق ذات الدلالة الإحصائية والتساؤلات البحثية الأخرى، ويوضح الجدول(٢٥) هذه التكرارات ومتوسطاتها على النحو التالي:

جدول ٢٥ : مجموع تكرارات التساؤلات عن الفروق ذات الدلالة الإحصائية وغيرها من التساؤلات

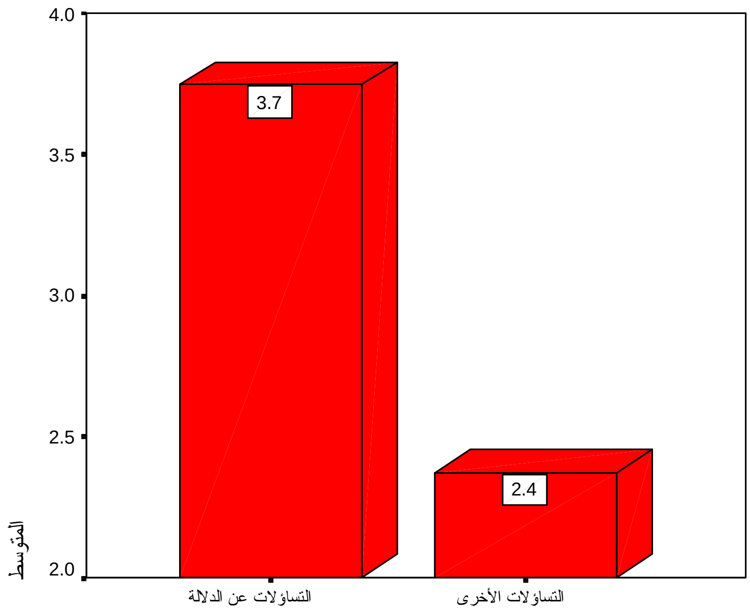
الحـــالـــه	التكرارات	المتوسط لكل رسالة
التساؤلات البحثية عن الفروق ذات الدلالة الإحصائية	٣٧١	٣,٧٥
التساؤلات البحثية الأخرى	٢٣٥	٢,٣٧
المجموع	٦٠٦	

يتضح من الجدول السابق أنّ عدد التساؤلات البحثية عن الفروق ذات الدلالة الإحصائية بلغ (٣٧١) تساؤلاً بحثياً مقارنة بـ (٢٣٥) تساؤلاً بحثياً عن إحصاءات أو مقارنات أخرى لا تخص اختبارات الدلالة الإحصائية.

الرسمان البيانيان التاليان يمثلان مجموع تكرارات التساؤلات عن الفروق الدالة إحصائياً ومجموع تكرارات غيرها من التساؤلات، وكذلك متوسطي التساؤلات عن الفروق الدالة إحصائياً وغيرها من التساؤلات:

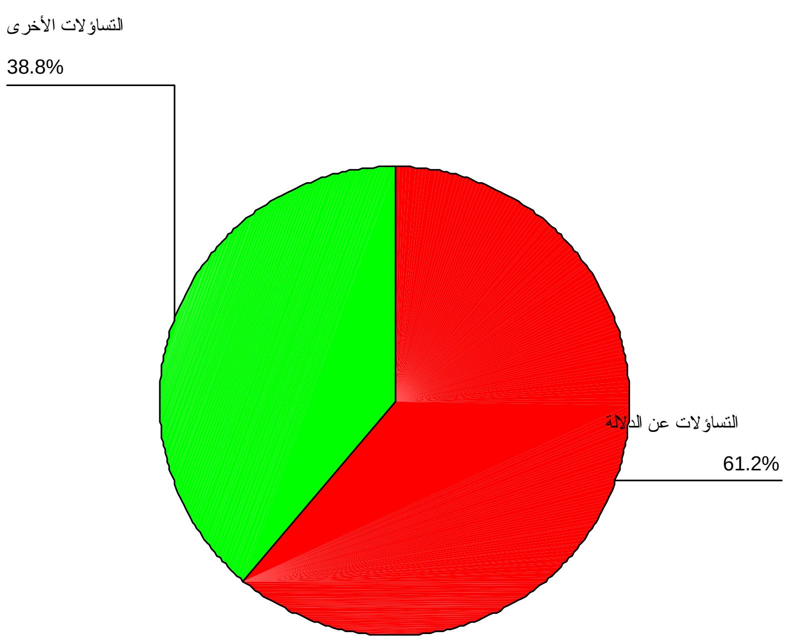


رسم بياني ٣٠ : تكرارات التساؤلات عن الفروق ذات الدلالة إحصائياً وغيرها من التساؤلات



رسم بياني ٣١ : متوسطا التساؤلات عن الفروق ذات الدلالة إحصائياً وغيرها من التساؤلات

الرسم البياني التالي يوضح النسبة المئوية للتساؤلات عن الفروق الدالة إحصائياً وغيرها من التساؤلات البحثية:



رسم بياني ٣٢ : النسبة المئوية للتساؤلات عن الفروق ذات الدلالة إحصائياً وغيرها من التساؤلات

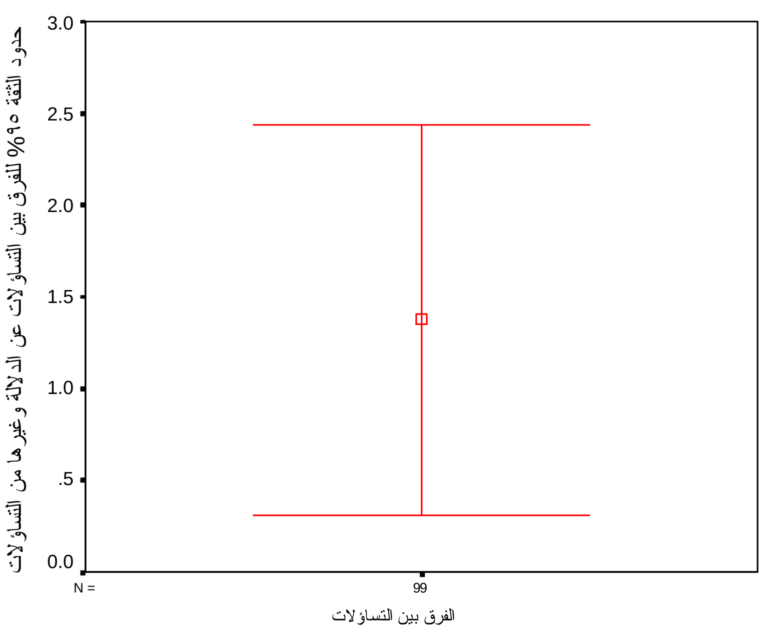
يتضح من الرسم البياني السابق أنّ عدد التساؤلات حول الفروق ذات الدلالة الإحصائية نسبتها (٦١,٢٪)، بينما التساؤلات البحثية الأخرى تشكل (٣٨,٨٪).

ولاختبار ما إذا كان الفرق بين متوسطي الأسئلة عن الفروق ذات الدلالة الإحصائية وبقية التساؤلات البحثية الأخرى ذات دلالة إحصائية أم لا، قام الباحث بحساب حدود الثقة للفرق بين المتوسطين حيث كانت النتيجة كما هو موضح في الجدول التالي:

جدول ٢٦ : حدود الثقة للفرق بين متوسطي التساؤلات عن الفروق ذات الدلالة الإحصائية وغيرها من التساؤلات

الحـــالـــه	الفرق بين المتوسطات	الحد الأعلى	الحد الأدنى
حدود الثقة عند مستوى ٩٥٪ للفرق بين متوسطي التساؤلات عن الفروق الدالة إحصائياً وغيرها من التساؤلات البحثية	١,٣٧	٠,٣١٠١	٢,٤٣٧٤

الرسم البياني التال يوضح تمثيلاً بيانياً لتقدير حدود الثقة للفرق بين متوسطي التساؤلات عن الفروق ذات الدلالة الإحصائية وغيرها من التساؤلات البحثية:



رسم بياني ٣٣ : تقدير حدود الثقة للفرق بين متوسطي التساؤلات عن الفروق ذات الدلالة إحصائياً وغيرها من التساؤلات

يتضح من الجدول والرسم البياني السابقين أنّ حدود الثقة عند مستوى ٩٥٪ للفرق بين متوسطي التساؤلات عن الفروق ذات الدلالة الإحصائية وغيرها من التساؤلات هي (٠,٣١ - ٢,٤٤). وحيث أنّ هذه الحدود لا تتضمن الصفر، فإنّ هذا يعني أنّ الفروق ذات دلالة إحصائية عند مستوى ٠,٠٥ لصالح التساؤلات عن الفروق ذات الدلالة الإحصائية.

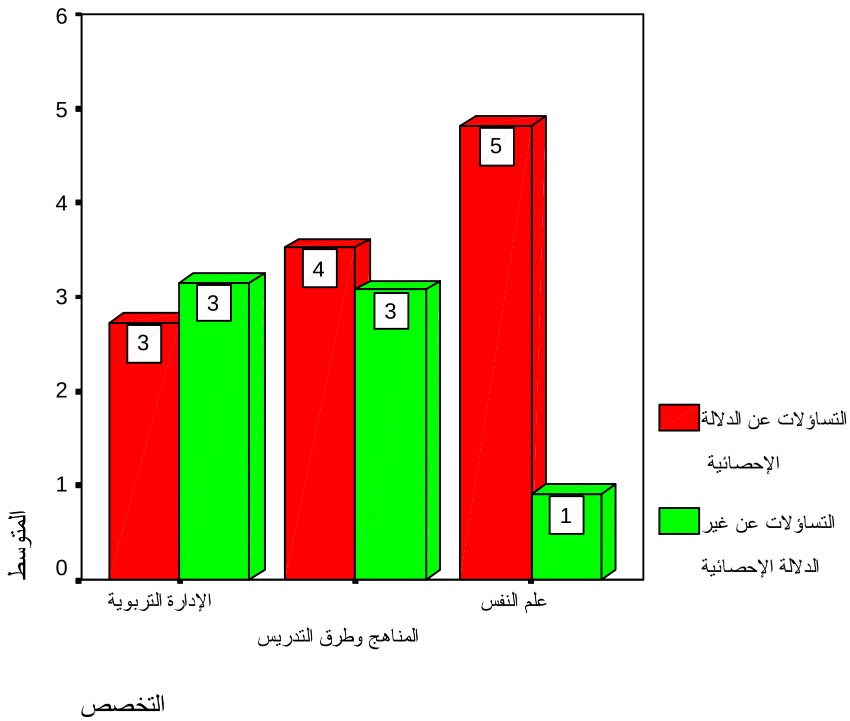
وهذا يعني وجود مشكلة في اهتمام رسائل الماجستير في كلية التربية بجامعة أم القرى بالتساؤلات عن الفروق ذات الدلالة الإحصائية والتي هي في الواقع لا تعني الكثير، بل أنّ الإجابة تقريباً معروفة مسبقاً حيث أنّ الفروق لابد من وجودها في العينات، وحتى إذا كانت الفروق ضئيلة فإنّ المسألة تعتمد على العدد الكافي من حجم العينة لإظهار تلك الفـروق ذات الدلالة الإحصائية(Snyder and Thompson,1998) . وتتفق نتائج هذه الدراسة مع ما ذكره سنايدر وثومبسون Snyder& Thompson(1998) وكيرك Kirk(1996) بأنّ توجه البحث التربوي حالياً هو للتساؤلات عن الفروق ذات الدلالة الإحصائية والتي تعتمد في إجابتها على رفض الفرضيات الصفرية.

الجدول التالي يوضح الفروق بين التساؤلات عن الفروق ذات الدلالة الإحصائية وغيرها من التساؤلات البحثية في أقسام كلية التربية الثلاثة:

جدول ٢٧ : مقارنة متوسطات عدد التساؤلات عن الفروق ذات الدلالة الإحصائية وغيرها من التساؤلات حسب أقسام كلية التربية

الحـــالـــه	الإدارة التربوية	المناهج	علم النفس
متوسط عدد التساؤلات عن الفروق ذات الدلالة الإحصائية	٢,٧٣	٣,٥٢	٤,٨٢
متوسط عدد التساؤلات البحثية الأخرى	٣,١٥	٣,٠٧	٠,٩١

الرسم البياني التالي يوضح تمثيلاً للفروق بين متوسطات عدد التساؤلات عن الفروق ذات الدلالة الإحصائية وغيرها من التساؤلات البحثية الأخرى حسب أقسام كلية التربية:



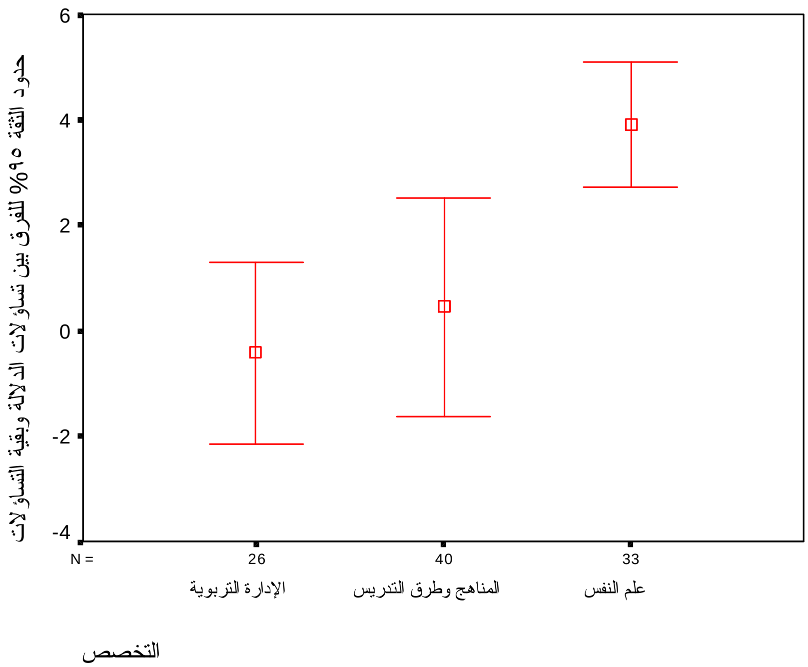
رسم بياني ٣٤ : متوسطات التساؤلات عن الفروق ذات الدلالة إحصائياً وغيرها من التساؤلات حسب أقسام كلية التربية

الجدول التالي يوضح الفرق بين متوسطات التساؤلات عن الفروق ذات الدلالة الإحصائية وغيرها من التساؤلات البحثية، وكذلك الحد الأعلى والأدنى لحدود الثقة عند مستوى ثقة ٩٥٪ لكل قسم من أقسام كلية التربية:

جدول ٢٨ : الفروق بين متوسطات التساؤلات عن الدلالة الإحصائية وغيرها من التساؤلات حسب أقسام كلية التربية

الحـــالــــه	الإدارة التربوية	المناهج	علم النفس
الفرق بين متوسطي التساؤلات عن الدلالة الإحصائية وغيرها من التساؤلات	-٠,٤٢	٠,٤٥	٣,٩١
الحد الأعلى لحدود الثقة عند مستوى ٩٥٪	١,٢٩	٢,٥٣	٥,٠٩
الحد الأدنى لحدود الثقة عند مستوى ٩٥٪	-٢,١٤	-١,٦٣	٢,٧٣

الرسم البياني التالي يوضح تمثيلاً لتقدير حدود الثقة للفرق بين المتوسطات في كل قسم من أقسام كلية التربية:



رسم بياني ٣٥ : حدود الثقة للفرق بين متوسطات التساؤلات عن الفروق ذات الدلالة إحصائياً وغيرها من التساؤلات حسب الأقسام

يتضح من الجدول والرسم البياني السابقين أنّ أعلى فرق بين التساؤلات عن الفروق ذات الدلالة الإحصائية وبقية التساؤلات البحثية هي في قسم علم النفس ويوضح الرسم البياني (٣٢) أنّ ذلك الفرق دال إحصائياً عند مستوى ثقة ٩٥٪ حيث أنّ حدود الثقة لم تتضمن الصفر، أمّا في قسم المناهج وطرق التدريس وقسم الإدارة التربوية والتخطيط فقد كانت الفروق بين المتوسطات غير دالة إحصائياً عند مستوى ثقة ٩٥٪، ممّا يعني عدم وجود فروق ذات دلالة إحصائية عند مستوى ٠،٠٥٪ بين متوسطي التساؤلات عن الفروق الدالة إحصائياً والفروق غير الدالة إحصائياً في قسمي (المناهج وطرق التدريس) و (الإدارة والتخطيط).

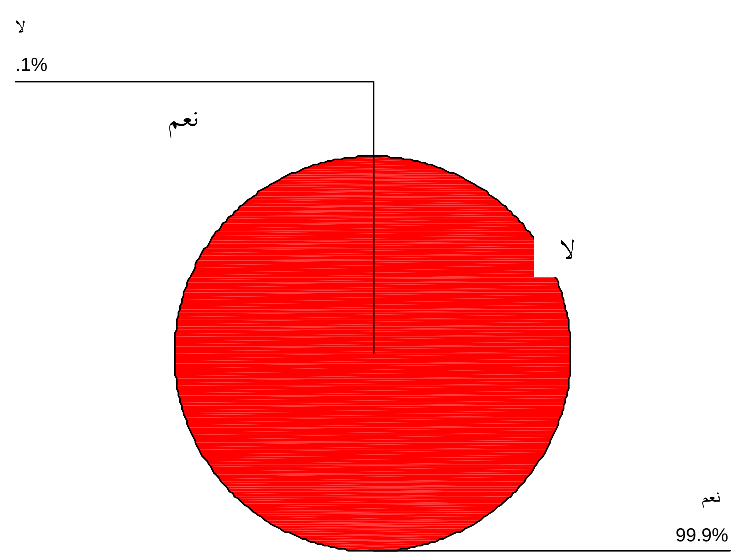
إجابة السؤال (٤ – ٨): ما مدى تأثر نتائج اختبارات الدلالة الإحصائية في رسائل الماجستير بكلية التربية بجامعة أم القرى بمأزق الطبيعة الثنائية للدلالة الإحصائية ؟

للإجابة عن هذا التساؤل قام الباحث بحصر نتائج اختبارات الدلالة الإحصائية في عينة الدراسة والتي كانت قيمة (r) فيها تتراوح من (٠،٠٥١) إلى (٠،٠٦). الجدول التالي يوضح عدد تكرارات النتائج التي تحمل القيمة من (٠،٠٥١) إلى (٠،٠٦) :

جدول ٢٩ : تكرارات الرسائل التي تتضمن قيم (**p**) ما بين (٠،٠٥١) و (٠،٠٦)

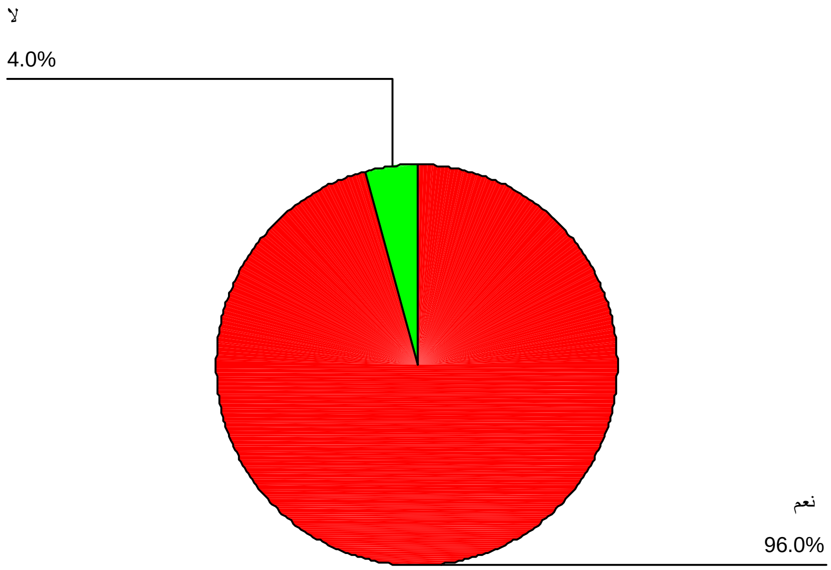
الحـــالـــة	عدد التكرارات		
	لم ترد مطلقاً	وردت مرّة واحدة	وردت مرّتين
الرسائل التي توجد بها قيم لـ (r) ما بين (٠،٥١) و (٠،٠٦)	٩٥	٣	١

يتضح من الجدول السابق أنّ (٩٥) رسالة لم ترد فيها هذه القيمة (٠،٠٥١ – ٠،٠٦)، بينما (٣) رسائل وردت فيها تلك القيمة مرّة واحدة في كل رسالة، بينما رسالة واحدة وردت فيها تلك القيمة مرتين، وبذلك يكون مجموع ورود النتيجة التي قيمة (r) فيها ما بين (٠،٠٥١ – ٠،٠٦) خمس مرّات من المجموع الكلي لنتائج اختبارات الدلالة الإحصائية البالغة (٥٤١٣) نتيجة، أي بنسبة ٠،١٪ من المجموع الكلي لنتائج اختبارات الدلالة الإحصائية، كما هو موضح في الرسم البياني التالي:



رسم بياني ٣٦ : : النسبة المئوية لتكرارات النتائج التي قيمة (**p**) فيها ما بين (٠،٠٥١) و (٠،٠٦) من التكرارات الكلية للنتائج

أمّا نسبة عدد الرسائل التي تتضمن قيمة (r) (٠،٠٥١ – ٠،٠٦) من العدد الكلي لرسائل العينة (٩٩) رسالة، فيوضحها الرسم البياني التالي:



الرسائل التي تحتوي على قيم الاحتمالية المحسوبة

ما بين ٠,٠٥١ و ٠,٠٦

رسم بياني ٣٧ : النسبة المئوية للرسائل التي تضمنت قيمة (**p**) فيها ما بين (٠,٠٥١) و (٠,٠٦) من العدد الكلي لرسائل العينة

يتضح من الرسم البياني السابق أنّ النسبة المئوية لعدد الرسائل التي تضمنت قيم (r) ما بين (٠,٠٥١) إلى (٠,٠٦) هي ٤,٠٪ من المجموع الكلي للرسائل في عينة الدراسة.

ممّا سبق يتضح أنّ نسبة النتائج التي تعاني من المأزق صغيرة جداً لا تتجاوز (٠,١٪) من إجمالي نتائج اختبارات الدلالة الإحصائية في عينة الدراسة، ولا تتجاوز نسبة الرسائل التي تتضمن مثل تلك القيم (٤,٠٪) من العدد الإجمالي للرسائل في عينة هذه الدراسة، وهذا يشير إلى أنّ هذه النسبة ليست كبيرة من حيث الكمية، ولكن من حيث المبدأ وحسب أهمية القرارات المطلوب اتخاذها فإنّ احتمالية حدوث مثل تلك النتيجة وارد جداً ممّا تتطلب معها مزيداً من الحذر في تفسير النتيجة واتخاذ القرار.

إجابة السؤال الخامس

إجابة السؤال (٥ – ١): ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى لتقدير حدود الثقة في تفسير نتائج اختبارات الدلالة الإحصائية؟

للإجابة عن هذا التساؤل قام الباحث بحساب التكرارات والنسبة المئوية لرسائل الماجستير التي استخدمت تقدير حدود الثقة في تفسير نتائج اختبارات الدلالة الإحصائية، ويوضح الجدول (٣٠) تلك التكرارات والنسبة المئوية على النحو التالي:

جدول ٣٠: التكرارات والنسبة المئوية لرسائل الماجستير التي استخدمت حدود الثقة في تفسير نتائج اختبارات الدلالة الإحصائية

الحـــالـــه	التكرارات	النسبة المئوية
الرسائل التي استخدمت حدود الثقة في تفسير النتائج	٠	صفر ٪
الرسائل التي لم تستخدم حدود الثقة في تفسير النتائج	٩٩	١٠٠٪
المجموع	٩٩	١٠٠٪

يتضح من الجدول السابق أنّ عدد الرسائل العلمية التي استخدمت حدود الثقة في تفسير نتائج الدلالة الإحصائية معدومة تماماً حيث بلغ عددها (صفراً). وتتفق هذه النتيجة تماماً مع ما توصل إليه كيرك Kirk(1996) بأنّه من المدهش أنّ امتياز وفائدة طريقة حدود الثقة لم تلق اهتماماً من قبل مجتمع الباحثين وأنّه قد تم تجاهلها، كما أنّها تختلف عمّا توصل إليه نيكس وبارنيت Nix and Barnette(1998) بأنّه يجب على الباحث أن يصف في دراسته حدود الثقة بشكل روتيني.

إجابة السؤال (٥ – ٢): ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى لتحليل القوة الإحصائية في تفسير نتائج اختبارات الدلالة الإحصائية؟

للإجابة عن هذا التساؤل قام الباحث بحساب التكرارات والنسبة المئوية لرسائل الماجستير التي استخدمت تحليل القوة الإحصائية في تفسير نتائج اختبارات الدلالة الإحصائية، ويوضح الجدول (٣١) تلك التكرارات والنسبة المئوية على النحو التالي:

جدول ٣١: التكرارات والنسبة المئوية لرسائل الماجستير التي استخدمت القوة الإحصائية في تفسير نتائج اختبارات الدلالة الإحصائية

الحـــالـــه	التكرارات	النسبة المئوية
الرسائل التي استخدمت حدود تحليل القوة في تفسير النتائج	٠	صفر ٪
الرسائل التي لم تستخدم تحليل القوة في تفسير النتائج	٩٩	١٠٠٪
المجموع	٩٩	١٠٠٪

يتضح من الجدول السابق أنّ عدد الرسائل العلمية التي استخدمت تحليل القوة الإحصائية في تفسير نتائج الدلالة الإحصائية معدومة تماماً حيث بلغ عددها (صفراً)، تتفق هذه النتيجة مع ما قرّره كل من كوهين Cohen(1988)، وشيفر (Wilkinson,1992) بأنّ التوجه لاختبار القوة الإحصائية نادراً ما يستخدم من قبل الباحثين.

ويرى الباحث أنّ هناك فرقاً مهماً يجب ملاحظته بين عدم اهتمام البحث الأجنبي ورسائل الماجستير بكلية التربية بجامعة أم القرى وهي أنّ في البحث الأجنبي عادة ما تنشر الدراسات التي وصلت إلى دلالة إحصائية لأنّ سياسة النشر تتحيّز إلى ذلك النوع من النتائج، وإذا حققت أي دراسة الدلالة الإحصائية فإنّ تحليل القوة الإحصائية يصبح قليل الجدوى لأنّه أصلاً يمنح الباحث تقديراً لاحتمالية رفض الفرضية الصفرية أي احتمالية الوصول إلى دلالة إحصائية. لكنّ الوضع في رسائل الماجستير بكلية التربية بجامعة أم القرى مختلف لأنّ نسبة النتائج غير الدالة إحصائياً (٥٢,٤٪) أعلى من نسبة النتائج غير الدالة إحصائياً، وهذا يعني أنّ الحاجة أشد لتحليل القوة الإحصائية لبرهنة أنّ تلك النتائج لم تكن بسبب انخفاض في القوة الإحصائية للاختبارات المستخدمة.

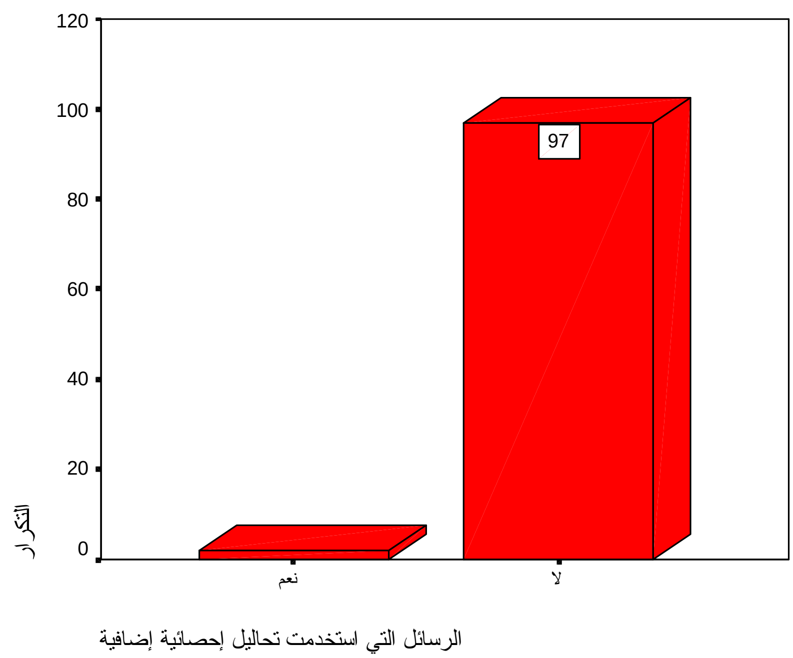
<u>إجابة السؤال (٥ – ٣): ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى مقاييس حجم التأثير كتحاليل متابعة لنتائج اختبارات الدلالة الإحصائية؟</u>

للإجابة عن هذا التساؤل قام الباحث بحساب التكرارات والنسبة المئوية لرسائل الماجستير التي استخدمت مقاييس حجم التأثير كتحاليل متابعة لنتائج اختبارات الدلالة الإحصائية، ويوضح الجدول (٣٢) تلك التكرارات والنسبة المئوية على النحو التالي:

جدول ٣٢: التكرارات والنسبة المئوية لرسائل الماجستير التي استخدمت حجم التأثير كمتابعة لنتائج اختبارات الدلالة الإحصائية

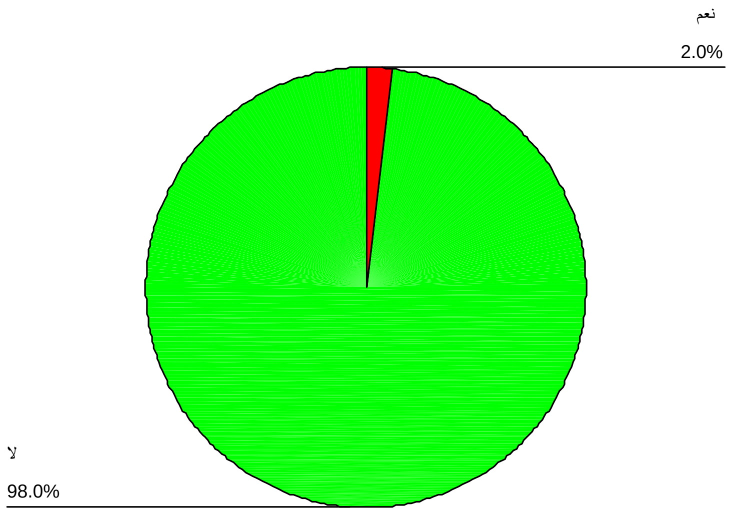
الحـــالـــه	التكرارات	النسبة المئوية
الرسائل التي استخدمت مؤشرات حجم التأثير مع اختبارات الدلالة	٢	٪ ٢,٠
الرسائل التي لم تستخدم مؤشرات حجم التأثير مع اختبارات الدلالة	٩٧	٪٩٨,٠
المجموع	٩٩	٪١٠٠

من الجدول السابق يتضح أنّ عدد الدراسات التي استعملت اختبارات متابعة إضافية لنتائج اختبارات الدلالة الإحصائية هي دراستين فقط، ويمثّل الرسم البياني التالي هذه المشاهدة:



رسم بياني ٣٨ : تكرارات رسائل الماجستير التي استخدمت أو لم تستخدم اختبارات متابعة

الرسم البياني التالي يمثّل النسبة المئوية للرسائل التي استخدمت اختبارات متابعة إضافية لاختبارات الدلالة الإحصائية:



رسم بياني ٣٩ : النسبة المئوية لرسائل الماجستير التي استخدمت أو لم تستخدم اختبارات متابعة

يتضح من الرسم البياني السابق أنّ الرسالتين اللتين اتبعتا اختبارات الدلالة الإحصائية باختبارات إضافية لا تمثّل إلاّ (٢٪) من مجموع رسائل عينة الدراسة. وتختلف نتيجة هذه الدراسة في مقدار النسبة المئوية عن نتائج بعض الدراسات السابقة، حيث بلغت نسبة الدراسات التي لم تستخدم تقديرات حجم التأثير في دراسة ثومبسون (Thompson(1999 ١٣٪ ، بينما كانت النسبة في دراسة ثومبسون وسنايدر (Thompson & Snyder(1997 ٣٦٪، أمّا في دراسة ماكلين (McLean(1995 والتي اختصت بالدراسات التي استخدمت اختباري t و F فقط فقد كانت النسبة فيها ٩٪ ، بينما في دراسة كيرك (Kirk(1996 فقد كانت النسبة ٢٣٪ .

ومهما اختلفت النسب المئوية للدراسات السابقة التي لم تستخدم أي مؤشر للدلالة العملية (حجم التأثير) كتقديرات مكملة لنتائج اختبارات الدلالة الإحصائية، فإنّ نسبة هذه الدراسة والتي تمثّل واقع رسائل الماجستير في كلية التربية بجامعة أم القرى تبقى الأقل وبفارق واضح.

وللتعرّف عن أنواع تلك الاختبارات الإضافية المستخدمة مع نتائج الدلالة الإحصائية في رسائل الماجستير بكلية التربية بجامعة أم القرى، قام الباحث بحصر تلك الأنواع وحساب تكراراتها كما يوضحها الجدول التالي:

جدول ٣٣: أنواع تقديرات حجم التأثير وتكراراتها والتي استخدمت كاختبارات مكملة لاختبارات الدلالة الإحصائية

الرسائل التي استخدمت مؤشرات الدلالة العملية	التكرارات	النسبة المئوية
r^2	١	٪ ١,٠
حجم التأثير	١	٪١,٠
لا شيء	٩٧	٪٩٨,٠
المجموع	٩٩	٪١٠٠

يلاحظ من الجدول السابق أنّ تقديرات الدلالة العملية التي استخدمت مع نتائج اختبارات الدلالة الإحصائية هي (r2) حيث استخدمت في رسالة واحدة وبالتحديد مرّتين في تلك الرسالة، وتقدير حجم التأثير حيث استخدم في رسالة واحدة أيضا وبالتحديد أربع مرّات في تلك الرسالة.

إجابة السؤال (٥ – ٤): ما مدى استخدام رسائل الماجستير في كلية التربية بجامعة أم القرى لتحليل إعادة النتائج؟

للإجابة عن هذا التساؤل قام الباحث بحساب التكرارات والنسبة المئوية لرسائل الماجستير التي استخدمت تحليل الإعادة، ويوضح الجدول (٣٢) تلك التكرارات والنسبة المئوية على النحو التالي:

جدول ٣٤: التكرارات والنسبة المئوية لرسائل الماجستير التي استخدمت تحليل الإعادة

الحـــالـــه	التكرارات	النسبة المئوية
الرسائل التي استخدمت تحليل الإعادة	٠	صفر ٪
الرسائل التي لم تستخدم تحليل الإعادة	٩٩	٪١٠٠
المجموع	٩٩	٪١٠٠

يتضح من الجدول السابق أنّ عدد الرسائل العلمية بكلية التربية بجامعة أم القرى لم تستخدم أي واحدة منها اختبار إعادة. وتتفق هذه النتيجة مع ما وجده ثومبسون وسنايدر (١٩٩٧) من أنّ عدد الدراسات التي استخدمت تحاليل الإعادة الداخلية في عينة دراستهم قد بلغت صفراً.

الفصل الخامس

الخلاصة والتوصيات

هـدفت هذه الدراسـة إلى:

- الكشف عن أهم مشكلات الدلالة الإحصائية في البحث التربوي.
- التعرّف على أهم المفاهيم الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في البحث التربوي.
- صياغة أساليب جديدة لحل مشكلات الدلالة الإحصائية في البحث التربوي.
- الكشف عن واقع مشكلات الدلالة الإحصائية في الرسائل العلمية لكلية التربية بجامعة أم القرى بمكة المكرمة.
- الكشف عن مدى توفر المفاهيم الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في الرسائل العلمية بكلية التربية بجامعة أم القرى بمكة المكرمة.

ولتحقيق تلك الأهداف السابقة حاولت الدراسة الإجابة عن التسـاؤلات التاليـة:

٦ـ ما أهم مشكلات الدلالة الإحصائية في البحث التربوي ؟

٧ـ ما أهم المفاهيم الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في البحث التربوي؟

٨ـ ما الأسـاليب الجديدة لحل مشكلات الدلالة الإحصائية في البحث التربوي؟

٩ـ ما واقع مشكلات الدلالة الإحصائية في الرسائل العلمية بكلية التربية بجامعة أم القرى بمكة المكرمة.

١٠ـ ما مدى وجود المفاهيم والأساليب الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في الرسائل العلمية لكلية التربية بجامعة أم القرى بمكة المكرمة.

وللإجابة عن التساؤلات السابقة اعتمد الباحث منهجين بحثيين، الأول يقوم على تحليل المعلومات حيث أجاب الباحث عن التساؤل الأول والثاني والثالث عن طريق البحث النظري الأساسي، والمنهج الثاني هو الوصفي التقويمي حيث أجاب الباحث على التساؤلين الرابع والخامس .

وللإجابة عن التساؤلين الرابع والخامس، قام الباحث باختيار عينة قصدية من رسائل الماجستير بكلية التربية بجامعة أم القرى التي أجيزت خلال الأعوام ١٤١٧، ١٤١٨، ١٤١٩ هـ التي استخدمت اختبارات الدلالة الإحصائية، حيث بلغ عدد الرسائل التي تم اختيارها من

أقسام (الإدارة التربوية والتخطيط، والمناهج وطرق التدريس، وعلم النفس) (٩٩) رسالة وبواقع (٧٢٪) من المجتمع الأصلي. وقد قام الباحث بجمع البيانات من خلال استمارة خاصة أعدّها الباحث لهذا الغرض.

وعلى ضوء ما سبق فقد قسم الباحث دراسته إلى خمسة فصول هي:

- الفصل الأول: تناول مشكلة الدراسة، وأهدافها، وأهميتها، وحدودها، وأهم مصطلحاتها.
- الفصل الثاني: تناول الإطار النظري لمفهوم الدلالة الإحصائية، والدراسات السابقة.
- الفصل الثالث: تناول منهج الدراسة المستخدم، وأداتها، ومجتمعها الإحصائي، وعينة الدراسة، والأساليب الإحصائية المستخدمة.
- الفصل الرابع: تناول الإجابة عن التساؤل الأول حول أهم مشكلات الدلالة الإحصائية في البحث التربوي، والإجابة عن التساؤل الثاني حول تحديد أهم المفاهيم التي يمكن أن تقدم حلولاً، والإجابة عن التساؤل الثالث حول الأساليب التي يمكن أن تستخدم معها تلك المفاهيم لتحسين تفسير نتائج اختبارات الدلالة الإحصائية. كما تناول هذا الفصل استعراض وتحليل البيانات وتفسير النتائج ومقارنتها بالدراسات السابقة وذلك للإجابة عن التساؤلين الرابع والخامس في هذه الدراسة.
- الفصل الخامس: حيث عرض فيه الباحث ملخصاً لأهم نتائج الدراسة، وكذلك التوصيات والمقترحات.

وقد توصلت الدراسة لما يأتي:

<u>أولاً – أهم مشكلات اختبارات الدلالة الإحصائية في البحث التربوي هي:</u>

١ - استخدام نتائج الدلالة الإحصائية كتفسير لأهمية النتائج.

٢ - استخدام نتائج الدلالة الإحصائية كتفسير لتأثير المعالجة.

٣ - استخدام قيمة الاحتمالية (r) لتقدير حجم التأثير.

٤ - استخدام قيمة الاحتمالية (r) كتفسير لاحتمالية الفرضية الصفرية نفسها.

٥ - استخدام قيمة الاحتمالية (r) لتقويم إعادة النتيجة.

٦ - التحيّز للنتائج الدالة إحصائياً.

٧ - صياغة التساؤلات البحثية والفرضيات الصفرية.

٨ - الطبيعة الثنائية لنتائج اختبارات الدلالة الإحصائية.

ثانياً - أنّ أهم المفاهيم الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في البحث التربوي هي:

١ - تحليل القوة الإحصائية.

٢ - تقدير حدود الثقة.

٣ - تقديرات حجم التأثير.

٤ - تحليل الإعادة.

ثالثاً - أنّ الأساليب الجديدة لحل مشكلات الدلالة الإحصائية في البحث التربوي هي:

١ - إعادة صياغة الأسئلة البحثية والفرضيات بحيث تركز على حجم التأثير وقيمة النتائج بدلاً من مجرد وجود أو عدم وجود الفروق.

٢ - فحص البيانات بشكل موسع عن طريق:

(أ) تحديد حجم التأثير المطلوب الكشف عنه لتحديد حجم العينة المطلوبة، وكذلك أهمية النتيجة.

(ب) حساب مؤشرات الدلالة العملية (حجم التأثير) بشكل مرافق لكل قيمة احتمالية محسوبة (r) .

(ج) تفسير حجم التأثير المحسوب في سياق البحث اعتماداً على أهداف البحث وطبيعة موضوع الدراسة.

(د) تفسير أثر حجم العينة في سياق تفسير نتائج الدراسة.

(ه) استخدام تقديرات حدود الثقة وتفسيرها.

(و) اختبار القوة الإحصائية للدراسات التي لم تصل إلى نتائج دالة إحصائياً.

(ز) عدم اعتبار أنّ الفرضية أو النظرية مبرهنة حتى يتم التحقق من تحاليل الإعادة.

(ح) إذا لم يتمكن الباحث من التحقّق من الإعادة الخارجية للدراسة، ينتقل إلى تحاليل الإعادة الداخلية كبديل مناسب.

رابعاً – واقع مشكلات الدلالة الإحصائية في رسائل الماجستير بكلية التربية بجامعة أم القرى هي على النحو التالي:

١ - توجد مشكلة كبيرة في درجة استخدام رسائل الماجستير بكلية التربية بجامعة أم القرى كتفسير لأهمية النتائج حيث بلغت نسبة تلك الرسائل (٩٨٪) رسالة اعتمدت فقط على نتائج اختبارات الدلالة الإحصائية في تفسير النتائج واعتبرت ذلك غاية الدراسة ومنتهاها.

٢ - توجد مشكلة في درجة استخدام رسائل الماجستير بكلية التربية بجامعة أم القرى كتفسير لتأثير المعالجات أو المتغيرات المستقلة حيث بلغت نسبة تلك الرسائل (٨٥,٩٪) رسالة استخدمت نتائج اختبارات الدلالة الإحصائية في تفسير تأثير المعالجات أو العوامل المستقلة.

٣ - توجد مشكلة استخدام قيم الاحتمالية (r) في تقدير حجم التأثير بنسبة ضئيلة حيث لم تتجاوز النسبة ٧,١٪ .

٤ - لا توجد أي مشكلة لاستخدام قيم الاحتمالية(r) في تقدير احتمالية إعادة النتائج حيث بلغت نسبة الاستخدام لهذا الغرض صفر٪.

٥ - لا توجد أي مشكلة لاستخدام قيم الاحتمالية (r) في تقدير احتمالية حدوث الفرضية الصفرية نفسها حيث بلغت نسبة الاستخدام لهذا الغرض صفر٪ .

٦ - لا توجد مشكلة تحيّز للنتائج التي وصلت لمستوى الدلالة الإحصائية مقارنة بالنتائج التي لم تصل إلى مستوى الدلالة الإحصائية حيث لم تتجاوز نسبة النتائج الدالة إحصائياً (٤٧,٦٪) من العدد الإجمالي لنتائج اختبارات الدلالة الإحصائية، كما أنّ اختبار الدلالة الإحصائية يظهر عدم وجود فروق ذات دلالة إحصائية عند مستوى ثقة ٩٥٪.

٧ - توجد مشكلة في مدى اعتماد رسائل الماجستير بكلية التربية بجامعة أم القرى على التساؤلات البحثية عن الفروق ذات الدلالة الإحصائية حيث بلغت نسبة استخدامها (٦١,٢٪) من مجموع التساؤلات البحثية في رسائل الماجستير بكلية التربية بجامعة أم القرى.

٨ - لا توجد مشكلة لتأثر نتائج اختبارات الدلالة الإحصائية بالطبيعة الثنائية لنتائج اختبارات الدلالة الإحصائية حيث لم تتجاوز النتائج التي قيمتها ما بين ٠,٠٥١ و ٠,٠٦ نسبة (٠,١٪) من المجموع الكلي لتلك لنتائج اختبارات الدلالة الإحصائية.

خامساً – واقع المفاهيم الإحصائية التي يمكن أن تقدم حلولاً لمشكلات الدلالة الإحصائية في رسائل الماجستير بكلية التربية بجامعة أم القرى هي على النحو التالي:

١ - لا يوجد أي استخدام لتحليل القوة الإحصائية في رسائل الماجستير بكلية التربية بجامعة أم القرى حتى مع النتائج غير الدالة إحصائياً، حيث بلغت النسبة الكلية الاستخدام تحليل القوة الإحصائية في الرسائل صفر٪ .

٢ - لا يوجد أي استخدام لتقدير حدود الثقة في تفسير نتائج الدلالة الإحصائية سواء كانت دالة إحصائياً أو غير دالة، حيث كانت النسبة الكلية لاستخدامها صفر ٪ .

٣ - يوجد استخدام ضئيل جداً لمقاييس حجم التأثير في رسائل الماجستير بكلية التربية بجامعة أم القرى حيث لم تتجاوز نسبة الاستخدام ٢٪.

٤ - لا يوجد أي استخدام لتحليل إعادة النتائج في رسائل الماجستير بكلية التربية بجامعة أم القرى حيث كانت نسبة الاستخدام صفر٪.

التوصيات

بناء على ما توصل إليه الباحث من نتائج فإنّه يوصي بما يلي:

١ - منح المزيد من الاهتمام لبحث وعرض مشكلات الدلالة الإحصائية في البحث التربوي المحلّي والعربي، حيث أنّ هذا الموضوع لم ينل نفس الاهتمام في العالم العربي كما ناله ولا يزال يناله من قبل الباحثين في العالم الغربي.

٢ - إعادة النظر في مقرّرات الإحصاء الاستدلالي المقدمة لطلاب الدراسات العليا بكلية التربية بحيث يراعى فيها ما يلي:

أ) تركيز المقرّرات على فلسفة الأساليب الإحصائية ومفاهيمها وحدودها وافتراضاتها ومشكلاتها ممّا يساعد الطالب في تفسير نتائجها.

ب) توجيه المقرّرات نحو المزيد من أساليب تقدير المؤشرات مقارنة بأساليب اختبار الفرضيات.

ج) توسيع خطوات اختبار الفرضيات التقليدية بحيث تتضمن قياس مؤشرات الدلالة العملية في نفس السياق.

د) تضمين المقرّرات مساحة أوسع لتدريب الطالب على تفسير معنى النتائج ومناقشتها ومقارنتها وصياغة الاستنتاجات.

٣ - تطوير معايير تحكيم الرسائل العلمية بحيث تضمن بشكل إجرائي ومحدّد عدم السماح للاعتماد الوحيد على تفسير نتائج اختبارات الدلالة الإحصائية دون ارفاقها بمؤشرات وتفسيرات الدلالة العملية.

٤ - وضع معايير جديدة لقبول نتائج اختبارات الدلالة الإحصائية بحيث تشترط تفسير النتائج غير الدالة إحصائياً وبرهنتها بمؤشرات أخرى.

٥ - إعادة النظر في التقاليد العلمية المتبعة والتي ترفض أي دراسة علمية إذا لم تكن جديدة في موضوعها، وذلك بتشجيع أساليب البحث التي تعتمد على إعادة الدراسات السابقة وتفسير نتائجها ومقارنتها خاصة إذا كان موضوع الدراسة مهماً وخطيراً ويستدعي المزيد من الثقة لبرهنة النظريات المحلّية المهمة، ووفق معايير علمية دقيقة تضمن جدوى ذلك النوع من الدراسات التي يمكن أن تسهم فعلاً في التراكم المعرفي للدراسات العربية، وكذلك تعرّف الباحثين على هذا النوع من الأبحاث وطرق تنفيذها والتعامل معها وتقدير أهميتها.

٦ - تأسيس مركز متخصص للإستشارات البحثية والإحصائية في جامعة أم القرى وذلك لتحسين المناخ العلمي في البيئة الجامعية خاصة على مستوى الدراسات العليا، بحيث يقدّم المركز الإستشارات لطلاب الدراسات العليا في تصميم التجارب وتحليل النتائج وتفسيرها، وكذلك تقديم الدورات التأهيلية والمحاضرات التثقيفية والندوات العلمية والنشرات المتخصصة في الأساليب الإحصائية وذلك لدعم أعضاء هيئة التدريس غير المتخصصين وارشاد ومساعدة جميع طلاب الدراسات العليا، على أن يدعم المركز بقومات المراكز المتخصصة من هيئة إشرافية من قبل أساتذة متخصصين في الإحصاء النفسي، وهيئة إدارية وفنية مساعدة ومرافق وأجهزة وميزانيات. كما يمكن أن

يقوم المركز بالمساعدة في رصد مدى تطوّر أو انحدار استخدام الأساليب الإحصائية على مستوى الجامعة وذلك لحماية البيئة العلمية من الممارسات غير الصحيحة، بحيث يقوم المركز بنشر أبرز ملاحظاته وتوصياته وقضايا اختصاصاته بالأساليب المتاحة والتي منها تأسيس موقع علمي رصين على شبكة الإنترنت، التي تقدّم جميع إمكانيات العرض والتفاعل للمهتمين. ويمكن عرض نتائج هذه الدراسة والدراسات الإحصائية الأخرى التي تظهر ضعف الممارسات الإحصائية في الأبحاث العلمية المتمثلة بالرسائل الجامعية وذلك لكسب التأييد والدعم لإنشاء مثل هذا المركز.

المقترحات

يرى الباحث في نهاية هذه الدراسة أن يتقدّم بالمقترحات التالية:

١ - إجراء دراسة مشابهة تشمل رسائل الماجستير والدكتوراة في جامعات محلّية أخرى، وكذلك في الجامعات العربية لمقارنة أوضاع الممارسات الفعلية للأساليب الإحصائية في البحث العربي.

٢ - إعادة تطبيق هذه الدراسة على المجلات العلمية المتخصصة في الدول العربية لمقارنة مشكلات الدلالة الإحصائية في تلك الدراسات بما هو موجود في رسائل طلاب الدراسات العليا، وكذلك للمقارنة بين مشكلات الدلالة الإحصائية في البحث التربوي العربي والبحث التربوي الغربي.

الخاتمة

بعد هذا الإبحار في مشكلات الدلالة الإحصائية ووسائل تحسين الممارسة، يستطيع الباحث أن يكون أكثر قدرة وموضوعية في وصف جوانب القوة وجوانب الضعف في هذه الدراسة. تكمن قوة هذا الدراسة كما يراها الباحث في الموضوع ذاته وما له من تطبيقات وتأثيرات على واقع ومستقبل البحث التربوي بشكل عام وفي جامعة أم القرى بشكل خاص. ويأمل الباحث أن تكون معالجته للموضوع وما توصل إليه من نتائج تتناسب مع خطورته وأهميته.

ويجد الباحث أنّ الأمانة العلمية وأخلاقيات البحث تقتضي أن يشير إلى مكامن الصعوبات والمشكلات التي واجهت هذه الدراسة. وفي رأي الباحث أنّ الصعوبات والمشكلات

التي واجهت هذه الدراسة اعتمدت على اختيار الباحث لعينة الدراسة، حيث يرى الباحث أنّ اختيار العينة من رسائل الماجستير بكلية التربية بجامعة أم القرى أوجد التحفظات التالية:

١ - أنّ معدّي تلك الرسائل لا يعتبرون محترفين في البحث التربوي، ولذا فمن الصعب الحكم على واقع مشكلات الدلالة الإحصائية في البحث العربي أو حتّى المحلّي من خلال نتائج هذه الدراسة بسبب أنّ رسائل الماجستير لا تعد بحوثاً محترفة ومحكمّة بنفس المستوى الذي يمارس مع الدراسات والمقالات العلمية التي ينشرها الأكاديميون المحترفون في المجلات المتخصصة.

٢ - نظراً لأنّ المراجع العربية التي كتبت لمعالجة مشكلات الدلالة الإحصائية بشكل مباشر ومستقل قليلة جداً مقارنة بما كتب حديثاً عن هذه المشكلات في المراجع الأجنبية، ونظراً لأنّ تلك المراجع الأجنبية الحديثة والمتقدمة جداً كانت أغلب عيناتها مجلاّت علمية مرموقة ومحترفة جداً، فقد سببت هذه المفارقات عدّة صعوبات منها:

(أ) مشكلة في تطبيق أداة الدراسة التي بنيت على الخلفية السابقة، حيث أنّ أثناء تطبيق الأداة لاحظ الباحث أنّ كثيراً من البنود لا يوجد لها تطبيق فعلي في رسائل الماجستير، أي أنّ ممارساتها كانت أصفاراً.

(ب) ما سبق يفسّر ظهور بعض نتائج تحليل البيانات التي لم تعكس أي وجود لبعض المشكلات، حيث يرى الباحث أنّ تلك المفارقات التي ذكرها كانت وراء ظهور ذلك النوع من النتائج.

(ج) كما ظهرت أحياناً صعوبة في مقارنة بعض نتائج هذه الدراسة بما وجد في الدراسات السابقة المتقدّمة جداً.

وبناء على ما سبق يرى الباحث ضرورة التحفظ على مقارنة الممارسات الإحصائية في البحث العربي أو حتى البحث المحلّي بالممارسات في البحث الغربي، وذلك لما ذكره الباحث سابقاً من تأثيرات عينة هذه الدراسة على المدى الذي يمكن أن يصل إليه تفسير واستنتاج النتائج.

كما أنّ الباحث يتحفظ على أي استنتاج من هذه الدراسة يذهب إلى القول بعدم وجود أي جدوى لاختبارات الدلالة الإحصائية أو التقويض الكلّي للممارسات الحالية لاختبارات الدلالة الإحصائية في البحث التربوي، ويأمل الباحث أن تكون هذه الدراسة قد

حققت هدفها فعلاً وهي استكشاف المشكلات التي تمارس في البحث التربوي من خلال اختبارات الدلالة الإحصائية، واستكشاف الاقتراحات والحلول التي يمكن أن تكون بديلة أو مكملة لما لا تستطيع اختبارات الدلالة الإحصائية القيام به.

وختاماً.. يسأل الله تعالى الباحث القبول والتوفيق، وصلى الله على نبينا محمد وآله وصحبه وسلم، وآخر دعوانا أن الحمد لله رب العالمين.

المراجع

المراجع العربية

١ - بدر، أحمد (١٩٨٤م). أصول البحث العلمي ومناهجه. وكالة المطبوعات.

٢ - الخولي، محمد علي (١٩٩٤م). قاموس التربية انجليزي -عربي. بيروت: دار العلم للملايين.

٣ - الصائغ، ابتسام (١٤١٥هـ). الدلالة الإحصائية والدلالة العملية لاختبار (ت) و (ف) . رسالة ماجستير غير منشورة. كلية التربية بجامعة أم القرى.

٤ - الصياد، عبدالعاطي أحمد (١٩٨٨م). الدلالة العملية وحجم العينة المصاحبتين للدلالة الإحصائية لاختبار **t** في البحث التربوي والنفسي (دراسة تقويمية). الزقازيق، جامعة الزقازيق.

٥ - الصياد، عبدالعاطي أحمد (١٩٨٩م). جداول تحديد حجم العينة في البحث السلوكي. مجلة رابطة التربية الحديثة، القاهرة، العدد الأول، يناير.

٦ - عبيدات، ذوقان و عدس، عبدالرحمن و عبدالحق، كايد (١٩٩٩م). البحث العلمي مفهومه أدواته أساليبه. عمان: دار اسامة للنشر والتوزيع.

٧ - فان دالين، ديوبولد ب (١٩٦٩م). مناهج البحث في التربية وعلم النفس. ترجمة محمد نبيل نوفل وآخرون. القاهرة: مكتبة الانجلو المصرية.

٨ - القرشي، عبدالفتاح إبراهيم (١٤٢١هـ). تصميم البحوث في العلوم السلوكية. الكويت: دار القلم، الطبعة الأولى.

٩ - النجار، عبدالله عبدالرحمن (١٤١١هـ). دراسة تقويمية مقارنة للأساليب الإحصائية التي استخدمت في تحليل البيانات في رسائل الماجستير في كل من كلية التربية بجامعة أم القرى بمكة المكرمة وكلية التربية بجامعة الملك سعود بالرياض. رسالة ماجستير غير منشورة. كلية التربية بجامعة أم القرى.

١٠ -نور، رجاء محمد أحمد (١٤١٣هـ). تقويم استخدامات اختبار كاي تربيع في رسائل الماجستير بكلية التربية. رسالة ماجستير غير منشورة. كلية التربية بجامعة أم القرى.

المراجع الأجنبية

1- Bangert-Drowns, Robert L, & Runder, Lawrence M.(1991).**Meta-Analysis in Educational Research**.

2- Barnette, J. Jackson, and McLean, James E. (2000). **Use of the Significance Test as a Protection against Spuriously High Standardized Effect Sizes: Introduction of the Protected Effect Size.** Paper Presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).

3- Breunig, Nancy A.(1995).**Understanding the Sampling Distribution and Its Use in Testing Statistical Significance.** Paper Presented at the Annual Meeting of the Mid-South Educational Research Association (Biloxi, MS, November 1995).

4- Burdenski, Thomas (1999). **A Review of the Latest Literature on Whether Statistical Significance Tests Should be Banned**. Paper Presented at the Annual Meeting of the Southwest Educational Research Association (San Antonio, TX, January 21-23, 1999).

5- Carver, Ronal P.(1993). The Case against Statistical Significance Testing, Revisited. **Journal of Experimental Education**. V61 n4, P287-92, Sum 1993.

6- Chow, Siu L. (1998). Precise of Statistical Significance: Rationale, Validity, and Utility. **Behavioral and Brain Sciences.** V21, p169-239, 1998.

7- Chow, Siu L., & Oaks, Thousand (1998). Statistical Significance; Rationale, Validity, and Utility. **Journal of the American Association**. March 1998.

8- Cohen, Jacob (1988). **Statistical Power Analysis for the Behavioral Sciences.** New York: Academic Press.

9- Cohen, Jacob (1990). Things I Have Learned (So Far). **American Psychologist**. V45, n12, P1304-12, Dec 1990.

10- Cowles, Michael, & Davis, Caroline (1982).On the Origins of the .05 Level of Statistical Significance. **American Psychologist**.V37, n5,P533-58,May 1982.

11- Cromwell, Susan (2001). **An Introductory Summary of Various Effect Size Choices.** Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, February 1-3, 2001).

12- Daniel, Larry G. (1998a). Statistical Significance Testing: A Historical Overview of Misuse and Misinterpretation with Implications for the Editorial Policies of Educational Journals. **Research in the Schools**. V5, n2, P23-32, Fall 1998.

13- Daniel, Larry G.(1998b). **The Statistical Significance Controversy Is Definitely Not Over: A Rejoinder to Responses by Thompson, Knapp, and Levin**. Research in the Schools.V5 n2, P63-65, Fall 1998.

14- Daniel, Larry G.(1997). **Statistical Significance Testing in "Educational and Psychological Measurement" and Other Journals**. Paper Presented at the Annual Meeting of National Council on Measurement in Education (Chicago, IL, March 25-27,1997).

15- Daniel, Larry G.(1996). **Kerlinger's Research Myths**. ERIC/AE Digest Series. Nov 1996, ED410232.

16- Deegear, James (2001). **Recent Literature on Whether Statistical Significance Tests Should or Should Not Be Banned.** Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, February 1-3,2001).

17- Fagley, N. S., & McKinny, I. Jean (1983). Reviewer Bias for Statistically Significant Results: A Reexamination. **Journal of Counseling Psychology**. V30, n2, P298-300, Apr 1983.

18- Fan,Xitao ; Wang, Lin; & Thompson, Bruce (1999). **Effects of Sample Size, Estimation Methods, and Model Specification on Structural Equation Modeling Fit Indexes**. Structural Equation Modeling. V6 n1, P56-83, 1999.

19- Fish, Larry (1986). **The Importance of Invariance Procedures as against Tests of Statistical Significance**. Paper Presented at Annual Meeting of the Mid-South Educational Research Association (Memphis, TN, November 20, 1986).

20- Friedman, Herbert (1982). Simplified Determinations of Statistical Power, Magnitude of Effect and Research Sample Sizes. **Educational and Psychological Measurement**. V42 n2, P521-26, Sum 1982.

21- Furukawa, Toshi A. (1999). From Effect Size into Number Needed to Treat. The Lancet. V353, May 15, 1999.

22- Gill, Martin (1993).The Significance of "Significance". **Edinburgh Working Papers in Applied Linguistics.** V4, P63-80, 1993.

23- Gillaspy, James Arthur (1996). **Evaluating Result Replicability. Better Alternatives to Significance Tests**. Paper Presented at Annual Meeting of the Southeast Educational Research Association (New Orleans, LA, January 27, 1996).

24- Gorman, Bernard S; Primavera, Louis H.; & Allison, David B. (1995). POWPAL: A program for Estimating Effect Sizes, Statistical Power, and Sample Sizes. **Educational and Psychological Measurement**. V55, n5, pp. 773-776, October 1995.

25- Hanson, Marjorie, & Others (1987). Practical Significance in Special Education Research. **Journal of Special Education.** V20, n4, P401-08, 1986-87.

26- Hanson, Marjorie, & Others (1979). **Statistical Models and Practical Significance in Reading Research**. Paper Presented at Annual Meeting of the National Reading Conference (29th, San Antonio, TX, November29-December 1,1979).

27- Harcum, E. Rae (1989). The Highly Inappropriate Calibrations of Statistical Significance. **American Psychologist**. June 1989.

28- Hasse, Richard F., & Others (1982).How Significance Is a Significant Difference? Average Effect Size of Research in Counseling Psychology. **Journal of Counseling Psychology**. V29, n1, P58-65, Jan 1982.

29- Hazleton, Vincent, & Riley, Patricia (1981).Significance Tests and Decision in Communication Research: A New Method for Reducing Type II Error. **Communication Quarterly**.V30, n1, P13-17,Win 1981.

30- Helms, LuAnn Sherbeck (1999). **Basic Concepts in Classical Test Theory: Tests Aren't Reliable, the Nature of Alpha, and Reliability Generalization as a Meta-analytic Method**. Paper Presented at Annual Meeting of the Southwest Educational Research Association (San Antonio,TX,January 21-23,1999).

31- Henson, Robin K., and Smith, A. Delany (2000). State of the Art in Statistical Significance and Effect Size Reporting: A Review of the APA Task Force Report and Current Trends. **Journal of Research and Development in Education**. V33, n4, p285-96, Sum 2000.

32- Hetrick, Sam (1999). **A primer on Effect Sizes: What They Are and How To Compute Them**. Paper Presented at Annual Meeting of the Southwest Educational Research Association (San Antonio,TX,January 21-23,1999).

33- Hopkins, Will G.(1997). **A New View of Statistics.** [Online Book].

34- Hubbard, Raymond, and Ryan, Patricia A. (2000). The Historical Growth of Statistical Significance Testing in Psychology and Its Future Prospects. **Educational and Psychological Measurement**. V60, n5, p6661-81, Oct 2000.

35- Huberty, Carl J., & Lowman, Laureen L. (2000). Group Overlap As A bases For Effect Size. **Educational and Psychological Measurement**. V60, n4, pp. 543-563, August 2000.

36- Hummel, Thomas J, & Lichtenberg, James W.(1999). **Limits of Meta-Analysis as a Basis for Justifying Individual Counseling Interventions**. Paper Presented at Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23,1999).

37- Humphreys, Lloyd G., & Others (1993).Further Comments on Reliability and Power of Significance Tests and Reliability, Power, Functions, and Relations: A Reply to Humphreys. **Applied Psychological measurement**. V17, n1, P11-16, Mar 1993.

38- Huston, Holly L.(1993). **Meaningfulness, Statistical Significance, Effect Size, and Power Analysis: A General Discussion with Implications for MANOVA**. Paper Presented at Annual Meeting of the Mid-South Educational Research Association (2nd, New Orleans, LA, November 9-12, 1993).

39- Kanekar, Suresh (1990). Statistical Significance As A Continuum. **American Psychologist**. February 1990.

40- Kaufman, Alan S.(1998). Introduction to the Special Issue on Statistical Significance Testing. **Research in the Schools**.V5, n2, P1, Fall 1998.

41- Keaster, Richard D.(1998).**Statistical Significance Testing: From Routine to Ritual**. Paper Presented at Annual Meeting of the Mid-South Educational Research Association (Louisville, KY, November 9-11, 1998).

42- Kellow, J. Thomas (1998). Beyond Statistical Significant Tests: The Importance of Using Other Estimates of Treatment Effects To Interpret Evaluation Results. **American Journal of Evaluation**. V19, n1, P123-34, Win 1998.

43- Kier, Frederick J.(1997). **Ways To Explore the Replicability of Multivariate Results (Since Statistical Significance Does not)**. Paper Presented at Annual Meeting of the Southeast Educational Research Association (Austin, TX, January 23-25,1997).

44- King, Jason E. (1997). **Methods of Assessing Replicability in Canonical Correlation Analysis (CCA).** Paper Presented at Annual Meeting of the Southeast Educational Research Association (20th, Austin, TX, January 25, 1997).

45- Kirk, Roger E. (1996). Practical Significance: A concept Whose Time Has Come. **Educational and Psychological Measurement**. V56, n5, 746-759, October 1996.

46- Knapp, Thomas R.(1999). The Use of Tests of Statistical Significance. **Mid-Western Educational Researcher**. V12, n2, P2-5, Spr 1999.

47- Knapp, Thomas R.(1998).Comments on the Statistical Significance Testing Articles. **Research in the Schools**. V5, n2, P39-41, Fall 1998.

48- Kraemer, Helena Chmura (1983). Theory of Estimation and Testing of Effect Sizes: Use in Meta-Analysis. **Journal of Educational Statistics**. v8, n2, p93-101, Sum 1983.

49- Lachman, Sheldon J. (1993). Statistically Significant Differnce or Probable NonChance Difference. **American Psychologist.** October 1993.

50- Lane, David M. (2000). **HyperStat Online**.

51- Levin, Joel R, and Robinson, Daniel H. (2000). Rejoinder: Statistical Hypothesis Testing, Effect Size Estimation, and the Conclusion Coherence of Primary Research Studies. **Educational Researcher**. V29, n1, p34-36, Jan-Feb 2000.

52- Levin, Joel R. (1998). What if There Were No More Bickering about Statistical Significance Tests? **Research in the School**. V5 n2, P43-53, Fall 1998.

53- Levin, Joel R, & Others (1993). Statistical Significance Testing from Three Perspectives and Interpreting Statistical Significance and Nonsignificance and the Role of Statistics in Research. **Journal of Experimental Education**. V61, n4, P378-93, Sum 1993.

54- Magee, Kevin N, & Overall, John E. (1992).Formulae for Estimating Rater Reliability from the Significance of Treatment Effects. **Education and Psychological Measurement**. V52, n3, P545-48, Fall 1992.

55- Markel, William D. (1985).Statistical Significance: A Misunderstood Concept. **School Science and Mathematics**. V85, n5, P361-66, May-Jun 1985.

56- Marzano, Robert J. (1998). **Theory-Based Meta-Analysis of Research on Instruction**.

57- McCall, Robert B. (1980). **Fundamental Statistics for Psychology**. Third Edition,1980.

58- McClure, John, & Suen, Hoi K. (1994). Interpreting of Statistical Significance Testing: A Matter of Perspective. **Topics in Early Childhood Special Education**. V14, n1, P88-100, Spr 1994.

59- McLean, James E., O'Neal, Marcia R., and Barnette, J. Jackson (2000). **Are All Effect Sizes Created Equal?** Paper Presented at the Annual Meeting of the Mid-South Educational Research Association (28th , Bowling Green, KY, November 15-17, 2000).

60- McLean, James E., & Kaufman, Alan S.(1998a). Statistical Significance Testing. **Research in the Schools**. V5, n2, Fall 1998.

61- McLean, James, & Ernest, James M.(1998b).The Role of Statistical Significance Testing in Educational Research. **Research in the Schools**.V5, n2,P15-22,Fall 1998.

62- McLean, James E. & Ernest, James M. (1997). **Has Testing for Statistical Significance Outlived Its Usefulness?** Paper Presented at Annual Meeting of the Mid-South Educational Research Association (26th, Memphis, TN, November 12-14, 1997).

63- Menon, Rama (1993). Statistical Significance Testing Should Be Discontinued in Mathematics Education Research. **Mathematics Education Research Journal**. V5, n1, P4-18, Sep 1993.

64- Mittag, Kathleen C. and Thompson, Bruce (2000). A National Survey of AERA Members' Perceptions of Statistical Significance Tests and Other Statistical Issues. **Education Researcher**. V29, n4, p14-20, May 2000.

65- Mittag, Kathleen C.(1999). **A National Survey of AERA Members' Perceptions of the Nature and Meaning of Statistical Significance Tests**. Paper Presented at Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23,1999).

66- Moore, Jenny (1999). **Some Basic Concepts in Meta-Analysis**. Paper Presented at Annual Meeting of the Southeast Educational Research Association (San Antonio, TX, January 21-23,1999).

67- Morse, David T. (1999) . Minsize2: A Computer Program for Determining Effect Size and Minimum Sample Size for Statistical Significance for Univariate, Multivariate, and Nonparametric Tests. **Educational and Psychological Measurement**. V59, n3, pp. 518-531, June 1999.

68- Murray, Leigh W. & Dosser, David A.(1987). How Significant Is a Significant Difference? Problems With the Measurement of Magnitude of Effect. **Journal of Counseling Psychology**. v34, n1, p68-72, Jan 1987.

69- Nace, Barry J. & Bleakley, Thomas H.(1990). How Much Evidence Is Enough? **Trial**. August 1990.

70- Onwuegbuzie, Anthony J. (2000). **Effect Sizes in Qualitative Research**. Paper Presented at the Annual Meeting of the Association for the Advancement of Education Research (AAER) (Ponte Vedra, FL, November 2000).

71- Ottenbacher, Kenneth (1984).The Significance of Power and the Power of Significance: Recommendations for Occupational Therapy Research. **Occupational Therapy Journal of Research**. V4, n1, P37-50, Jan 1984.

72- Palomares, Ronald S.(1990). **Alternatives to Statistical Significance Testing**. Paper Presented at Annual Meeting of the Mid-South Educational Research Association (19th, New Orleans, LA, November 14-16, 1990).

73- Plucker, Jontahan A.(1997). Debunking the Myth of the "Highly Significant" Result: Effect Size in Gifted Education Research. **Roeper Review**. v20, n2, p122-26, Dec 1997.

74- Rafilson, Fred (1991). **The Case of Validity Generalization. Eric/TM Digest.**

75- Rennie, Kimberly (1997). **Understanding The Sampling Distribution: Why We Divide By n-1 To Estimate The Population Variance.** Paper presented at the annual meeting of the Southeast Educational Research Association. Austin, January 1997.

76- Rosenthal, Robert (1992). Effect Size Estimation, Significane Testing, and The Filedrawer Problem. **Journal of Parapsychology.** V56, 1992.

77- Rosnow, Ralph L. & Rosenthal, Robert (1988). Focused Tests of Significance and Effect Size Estimation in Counseling Psychology. **Journal of Counseling Psychology**. V35, n2, P203-08, Apr 1988.

78- Schmidt, Frank & Hunter, John E. (1995). The Impact of Data-Analysis Methods on Cumulative Research Knowledge: Statistical Significance Testing, Confidence Intervals, and Meta-Analysis. **Evaluation and Health Professions**. V18, n4, P408-27, Dec 1995.

79- Schmidt, Frank L.(1985). **From Validity Generalization to Meta-Analysis: The Development and Application of a New Research Integration Procedure**. Paper Presented at Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).

80- Schmidt, Frank L. & Hunter, John E.(1980). The Future of Criterion-Related Validity. **Personnel Psychology**. V33, n1, P41-57, Spr 1980.

81- Shaver, James P.(1992). **What Statistical Significance Testing Is, and What It Is Not**. Paper Presented at Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).

82- Shaver, James P.(1985a).Chance and Nonsense: A Conversation about Interpreting Tests of Statistical Significance, Part 1.**Phi Delta Kappan**. V67 n1, 57-60, Sep 1985.

83- Shaver, James P. (1985b). Chance and Nonsense: A Conversation about Interpreting Tests of Statistical Significance,Part2. **Phi Delta Kappan**. V67 n2, P138-41, Oct 1985.

84- Simon, Marilyn K.(1997). Statistical Significance: Rationale, Validity and Utility [book review]. **Canadian Journal of Program Evaluation**. V12, n2, P189-90, Aut 1997.

85- Slak, Marion K. (1997). Interpreting and Using Information from Studies without Statistical Significance. **The Annals of Pharmacotherapy.** v31, January 1997.

86- Smith, A. Delany, and Henson, Robin K.(2000). **State of the Art in Statistical Significance Testing: A Review of the APA Task Force on Statistical Inference's Report Current Trends.** Paper presented at the Annual Meeting of the Southwest Educational Research Association (Dallas,TX, January 27-29, 2000).

87- Snyder, Patricia & Thompson, Bruce.(1998). Use of Tests of Statistical Significance and Other Analytic Choices in a School Psychology Journal: Review of Practices and Suggested Alternatives. **School Psychology Quarterly**. V13, n4, P335-48, Win 1998.

88- Snyder, Patricia & Lawson, Stephen (1992). **Evaluating Statistical Significance Using Corrected and Uncorrected Magnitude of Effect Size Estimates**. Paper Presented at Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).

89- Spencer, Boyd (1995). Correlations, Sample Size, and Practical Significance: A Comparison of Selected Psychological and Medical Investigations. **The Journal of Psychology**. V129, n2, pp. 469-475, 1995.

90- StatSoft (2000). **Electronic Textbook.**

91- Stewart, David W. (2000). Testing Statistical Significance Testing: Some Observations of An Agnostic. **Educational and Psychological Measurement**. V60, n5, p685-690, October 2000.

92- Stockburger, David W.(1998). **Introductory Statistics: Concepts, Models, and Applications**. WWW Version 1.0, Revised 1998.

93- Strube, Michael J. (1988). Some Comments on the Use of Magnitude-of-Effect Estimates. **Journal of Counseling Psychology.** v35, n3, p342-45, Jul 1988.

94- Suen, Hoi K.(1992). Significance Testing; Necessary but Insufficient. **Topics in Early Childhood Special Education.** V12, n1, P66-81, Spr 1992.

95- Sullivan, Jeremy R.(2000). **A Review of Post-1994 Literature on Whether Statistical Significance Tests Should Be Banned.** Paper presented at the Annual Meeting of the Southwest Educational Research Association (Dallas,TX, January 27-29, 2000).

96- Thompson, Bruce (1999a). Improving Research Clarity and Usefulness with Effect Size Indices as Supplement to Statistical Significance Tests. **Exceptional Children**. V65, n3, pp. 329-337, 1999.

97- Thompson, Bruce (1999b). Why "Encouraging" Effect Size Reporting Is Not Working: The Etiology of Researcher Resistance to Changing Practices. **The Journal of Psychology**. V133, n2, pp. 133-140, 1999.

98- Thompson, Bruce (1999c). **Common Methodology Mistakes in Educational Research, Revisited, along with a Primer on Both Effect Sizes and the Bootstrap**. Paper Presented at Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23,1999).

99- Thompson, Bruce (1998a). Statistical Significance and Effect Size Reporting: Portrait of a Possible Future. **Research in the Schools**. V5, n2, P33-38, Fall 1998.

100- Thompson, Bruce (1998b). **Why "Encouraging" Effect Size Reporting Isn't Working: The Etiology of Researcher Resistance to Changing Practices**. Paper Presented at Annual Meeting of the Southwest Educational Research Association (Housten, TX, January, 1998).

101- Thompson, Bruce (1998c). **Five Methodology Errors in Educational Research: The Pantheon of Statistical Significance and Other Faux Pas**. Paper Presented at Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).

102- Thompson, Bruce & Snyder, Patricia A. (1998d). Statistical Significance and Reliability Analyses in Recent Journal of Counseling & Development Research Articles. **Journal of Counseling & Development**. V76,pp. 436-41,Fall 1998.

103- Thompson, Bruce (1997). Rejoinder: Editorial Policies Regarding Statistical Significance Tests: Future Comments. **Educational Research**.V26, n5,P29-32,Jun-Jul 1997.

104- Thompson, Bruce & Snyder, Patricia A. (1997a). **Use of Statistical Significance Tests and Reliability Analysis in Published Counseling Research**. Paper Presented at Annual Meeting of the American Educational Research Association (Chicago, IL, March 1997).

105- Thompson, Bruce & Synder, Patricia A.(1997b). Statistical Significance Testing Practices in “The Journal of Experimental Education.” **Journal of Experimental Education.** V66, n1, P75-83, Fall 1997.

106- Thompson, Bruce (1995a). **Inappropriate Statistical Practices in Counseling Research: Three Pointers for Readers of Research Literature**. ERIC Digest, 1995 , ED391990.

107- Thompson, Bruce (1995b). **Editorial Policies Regarding Statistical Significance Testing: Three Suggested Reforms**. Paper Presented at the Annual Meeting of the Mid-South Educational Research Association (Biloxi, MS, November 1995).

108- Thompson, Bruce (1994). **Common Methodology Mistakes in Dissertations, Revisited**. Paper Presented at Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-6,1994).

109- Thompson, Bruce (1988). **Common Methodology Mistakes in Dissertation Quality**. Paper Presented at Annual Meeting of the Mid-South Educational Research Association (Louisville, KY, November 8-11, 1988).

110- Trochim, William M.(1999). **Research Methods Knowledge Base**. 2nd edition,1999.

111- Vacha-Haase, Tammi & Nilson, Johana E.(1998).Statistical Significance Reporting: Current Trends and Uses in MECD. **Measurement and Evaluation in Counseling and Development**.V3,1 n1,P46-57, Apr 1998.

112- Vacha-Haase, Tammi & Thompson, Bruce (1998). Further Comments on Statistical Significance Tests. **Measurement and Evaluation in Counseling and Development**. V31, n1, P63-67, Apr 1998.

113- Wampold, Bruce E.(1983). Statistical Significance, Power, and Effect Size: A Response to the Reexamination of Review Bias. **Journal of Counseling Psychology**. V30, n3, P459-63, Jul 1983.

114- Welge-Crow, Patricia A, & Others (1990). **Looking Beyond Statistical Significance: Result Importance and Result Generalizability**. Paper Presented at Annual Meeting of the American Psychological Society (Dallas, TX, June 9,1990).

115- Westernmann, Rainer & Hager, Willi (1983). The Relative Importance of Low Significance Level and High Power in Multiple Tests of Significance. **Pe**

116- Wilkerson, Matt & Olson, Mary R. (1997). Misconceptions About Sample Size, Statistical Significance, and Treatment Effect. **The Journal of Psychology**. V131, n6, pp. 627-631, 1997.

117- Winer, Russell S. (2000). Comment on “The Historical Growth of Statistical Significance Testing in Psychology – and Its Future Prospects”. **Educational and Psychological Measurement**. V60, n5, pp. 693-696, October 2000.

118- Young, Martin A.(1993).Supplement Tests of Statistical Significance: Variation Accounted For. **Journal of Speech and Hearing Research**. V36, n4, P644-56, Aug 1993.

الملاحق

ملحق (١)

أداة الدراسة

(الصورة الأولية)

أداة الدراسة

م	رقم الرسالة في الدليل	عدد اختبارات الدلالة الإحصائية	عينة الدراسة		عدد استخدام مؤشرات إحصائية أخرى				اللغة العلمية المستعملة			تفسير ثبات النتائج				عدد استخدامات الدلالة الإحصائية في تقويم الافتراضات	التحيّز للدلالة الإحصائية		أهمية النتائج غير الدالة		المناسبة مع التساؤلات		شرح القيم المهدّدة للنتائج
			عددها	عدد تفسير أثرها على الدلالة	حدود الثقة	حجم التأثير	القوة الإحصائية	الدلالة العملية	دالة	دالة إحصائياً	مقارنة الدلالة	حساب ثبات الدرجات	اقتباس ثبات الاختبار	عدم الإشارة للثبات	استعمال P لتفسير الثبات		عدد الدالة	عدد غير الدالة	تفسير النتائج	عدد اختبارات المتابعة	عدد المناسبة	عدد غير المناسبة	

ملحق (٢)

أداة الدراسة

(الصورة النهائية)

أداة الدراسة

معلومات أساسية

رقم الرسالة: .... تاريخ النشر: ....

التخصص : £ إدارة تربوية وتخطيط £ مناهج وطرق تدريس £ علم نفس

الأساليب الاحصائية المستخدمة: £ اختبارات دلالة إحصائية £ إحصاء وصفي £ ليست احصائية

تساؤلات الدراسة وإجراءاتها

عدد التساؤلات عن الفروق أو الارتباطات ذات الدلالة الإحصائية	عدد التساؤلات الأخرى

اختبارات الدلالة الإحصائية المستخدمة في الدراسة

نوع اختبار الدلالة الاحصائية المستخدم	الدلالة الإحصائية	العدد	عدد ماتمّ متابعتها بتحاليل أخرى	عدد اختبارات المتابعة	أنواع اختبارات المتابعة المستخدمة
	دالة				
	غير دالة				
	دالة				
	غير دالة				
	دالة				
	غير دالة				
المجموع	دالة				
	غير دالة				

مشكلات الدلالة الإحصائية

	المشكله	نعم	لا	ملاحظات
	استخدام مؤشرات أخرى إضافة لاختبارات الدلالة الإحصائية			
	استخدام نتائج الدلالة الإحصائية كتفسير لتأثير المعالجة			
	استخدام عبارة (عالي الدلالة / مرتفعة الدلالة)			
	استخدام عبارة (تقترب من الدلالة / تقريباً دالة)			
	استخدام عبارات المقارنة (أعلى دلالة ، ...)			
	استخدام عبارة (دال / دالة) فقط.			
	استخدام r كاحتمالية للفرضية الصفرية نفسها			
	استخدام r لتقويم تكرار النتيجة			
	وجود قيم r تساوي (٠,٠٥١ – ٠,٠٦)			

الحلول البديلة

الأسلوب الإحصائي	عدد تكرارات استخدامه	عدد مرّات تفسيره	ملاحظات
القوة الإحصائية			
حدود الثقة			
تحليل تكرار النتيجة			

ملحق (٣)

أسماء السادة محكمي الأداة

أسماء محكمي أداة الدراسة

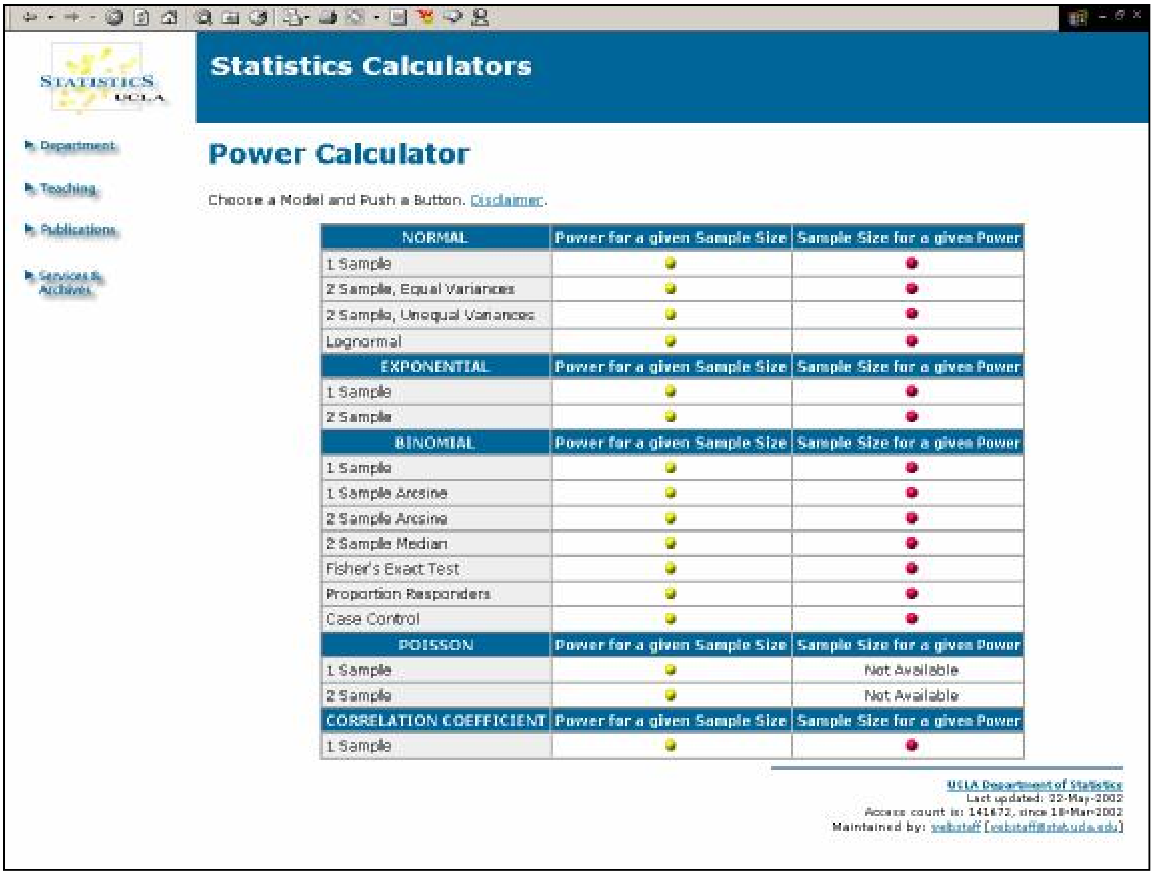
م	الاسم	الدرجة العلمية	جهة العمل
١	أ.د. ربيع بن سعيد طه	دكتوراة	قسم علم النفس بجامعة أم القرى
٢	أ.د. زايد بن عجير الحارثي	دكتوراة	قسم علم النفس بجامعة أم القرى
٣	د. ثابت بن محمد بن صالح القحطاني	دكتوراة	قسم علم النفس بجامعة أم القرى
٤	أ.د. عبدالحفيظ بن سعيد مقدم	دكتوراة	أكاديمية الأمير نايف للعلوم الأمنية
٥	د. عبدالله بن صالح المقبل	دكتوراة	وزارة المعارف
٦	د. عزت عبدالمجيد	دكتوراة	وزارة المعارف
٧	أ. عبدالله بن عمر بن جحلان	ماجستير	وزارة المعارف

ملحق (٤)

نموذج الصفحة التي يتم عن طريقها حساب القوة الإحصائية من موقع جامعة UCLA على شبكة الإنترنت

العنوان هو: http://www.stat.ucla.edu

حاسبة القوة الإحصائية كما تبدو من موقع جامعة UCLA



Statistics Calculators

Department

Teaching

Publications

Services & Archives

Power Calculator

Choose a Model and Push a Button. Disclaimer.

NORMAL	Power for a given Sample Size	Sample Size for a given Power
1 Sample		
2 Sample, Equal Variances		
2 Sample, Unequal Variances		
Lognormal		
EXPONENTIAL	Power for a given Sample Size	Sample Size for a given Power
1 Sample		
2 Sample		
BINOMIAL	Power for a given Sample Size	Sample Size for a given Power
1 Sample		
1 Sample Arcsine		
2 Sample Arcsine		
2 Sample Median		
Fisher's Exact Test		
Proportion Responders		
Case Control		
POISSON	Power for a given Sample Size	Sample Size for a given Power
1 Sample		Not Available
2 Sample		Not Available
CORRELATION COEFFICIENT	Power for a given Sample Size	Sample Size for a given Power
1 Sample		

UCLA Department of Statistics
Last updated: 22-May-2002
Access count is: 141672, since 18-Mar-2002
Maintained by: webstaff [webstaff@stat.ucla.edu]